Factors affecting speech reception in fluctuating noise and reverberation

George, E.L.J.

2007

Factors affecting speech reception in fluctuating noise and reverberation

The research described in this thesis was carried out within the EMGO Research Institute, at the ENT/Audiology Department of the VU University Medical Center in Amsterdam, The Netherlands.

This project was supported by the Heinsius-Houbolt Foundation, The Netherlands.

Cover:
In Search of Meaning #1
An Art and Science Collaboration
Artist - Lylie Fisher
http://www.lyliefisher.com

The cover displays a work of art inspired by bubble chamber experiments, which are used in particle physics to determine the trajectory of electrically charged particles. The shown patterns may be regarded as a visual metaphor for the complex mixture of sounds (speech, background noise and reverberation) that a hearing-impaired listener experiences in daily life.

ISBN: 978-90-9022332-2

Financial support for publication of this thesis was kindly provided by:
Stichting Atze Spoor Fonds, Beltone Nederland BV, Beter Horen BV, Oticon Nederland BV, HINK Groep BV, ReSound BV, Siemens Audiologie Techniek BV, Veenhuis Medical Audio BV

Printed by Gildeprint Drukkerijen BV, Enschede

VRIJE UNIVERSITEIT

FACTORS AFFECTING SPEECH RECEPTION IN FLUCTUATING NOISE AND REVERBERATION

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Geneeskunde
op vrijdag 16 november 2007 om 10.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Erwin Laura Johannes George

geboren te Heerlen

promotoren: prof.dr.ir. J.M. Festen
prof.dr.ir. T. Houtgast

Contents

General introduction

1

I. INTRODUCTION

A. Understanding speech

The ability to produce and listen to sophisticated speech is one of the main differences between humans and other biological species, and expressing yourself by saying what you mean is the major means of communication among humans. The complex processes that underlie speech communication are dazzling: it involves an incredibly ingenious cooperation between the human brain, vocal folds, mouth and ears. Even more amazing may be the fact that effective discourse among humans is possible. Apparently, people are not only able to produce and listen to speech; they are also able to understand, interpret and react to what is said by other talkers!

However, understanding speech can be a serious challenge in everyday life. Typical listening situations include a large variety of distortions that adversely affect speech intelligibility. Think about the noise caused by competing talkers or traffic, or the reverberation in large hallways in railway stations, churches or even hospitals. In addition to these external adverse factors, speech intelligibility can also be seriously hampered by internal, human factors. About 1.4 million people living in the Netherlands are hearing-impaired (TNS NIPO, 2005). They often complain of not being able to understand speech, even when the sound is presented at levels well above their detection threshold. Hearing-impaired listeners appear to be more vulnerable to the disturbing effect of noise and reverberation. When the background noise level fluctuates over time, the differences in speech reception between normal-hearing and hearing-impaired listeners appear to be even larger (Festen and Plomp, 1990). Understanding the processes behind this effect is important, since fluctuating backgrounds are very common in everyday situations (Kramer et al., 1996).

This introductory chapter gives an overview of the nature of speech processing, the disturbing influences of noise and reverberation, and the expected effect of hearing-impairment on speech intelligibility.

B. The nature of auditory speech processing

1. Speech as a physical sound

From a physical point of view, speech is a typical complex of air vibrations, in which sounds of various frequencies contribute at various times to produce a meaningful word or sentence. This acoustical content of speech can be visualized by means of a spectrogram, a two-dimensional figure with frequency along the vertical axis and time along the horizontal axis. The amplitude of the sound is displayed by shades of darkness. A spectrogram of a typical Dutch sentence is displayed in Fig. 1. A pure tone would be represented by a straight horizontal line, while a straight vertical line would represent a delta pulse, i.e. a click containing all frequencies. A typical sentence appears as a pattern of grey shades, which may seem rather disorganized at first. However, vowels in a sentence may be readily identified: their spectrum consists of various restricted frequency bands, which are caused by the resonance frequencies of the shape of the talker's mouth. The fundamental frequencies of the vocal folds and their harmonics form the horizontal lines in the shown spectrogram. As the pitch changes over time, these lines become sloping. In contrast, the energy of most consonants is noise-like in appearance, which is particularly clear for the fricatives (like 's' or 'v'), for which the energy is concentrated at the higher frequencies.

2. Speech in the auditory system

The air vibrations of speech sounds, as displayed in the spectrogram, are converted to nerve impulses by the human auditory system, consisting of the external, the middle and the internal ear. The external ear includes the auricle and the ear canal, which direct the sound to the tympanic membrane. The shape and length of the human external ear are such, that they prefer directing sound in the speech frequencies (Shaw and Teranishi, 1968). The tympanic membrane transduces the vibrations of the air to mechanical vibrations of the ossicles in the middle ear, suitable to stimulate the fluid filled inner ear or cochlea. The cochlea is a snail-shaped organ, filled entirely with a water-like liquid, which moves in response to the oscillations from the ossicles. Along its full length, the cochlea is divided in two by the basilar membrane, which 'houses' the organ of Corti with the auditory hair cells. The basilar membrane is set into motion by fluid pressure differences, which causes the hairs of the hair cells to vibrate. The hair cells convert the vibrations into nerve impulses, which travel along the auditory nerve to structures in the brainstem for further processing. Hair cells are distributed all the way along the basilar membrane. Each part of the basilar membrane, and the corresponding group of hair cells, is tuned to be responsive to a specific frequency. This well-structured tonotopic organization enables the auditory system to process frequencies from 20 Hz up to 20 kHz.

There is a limit to the temporal and spectral acuities with which changes in the time- or frequency-domain of a sound can still be followed by the auditory system. Auditory spectral resolution is mainly determined by the response characteristics of

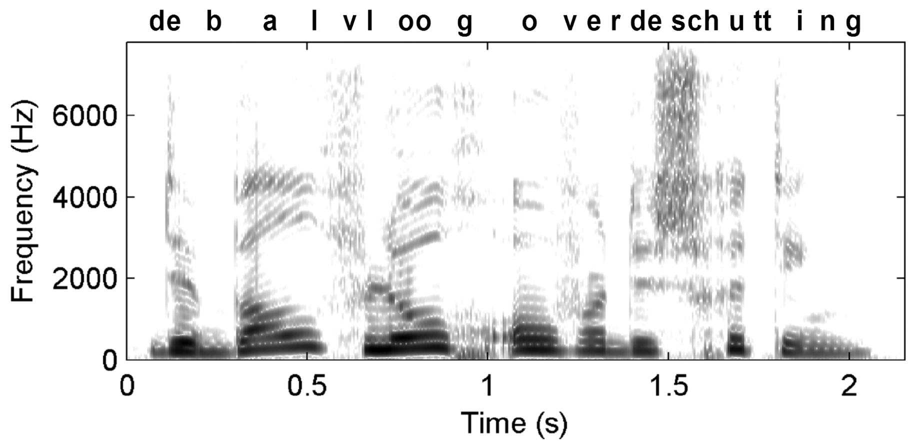

Figure 1-1 A spectrogram of a typical Dutch sentence: 'De bal vloog over de schutting' ['The ball flew over the fence']. Taken from the sentence lists as developed by Plomp and Mimpen (1979).

the hair cells. Fletcher (1940) suggested that the peripheral auditory system may be regarded as consisting of a bank of critical bandpass filters, currently known as the auditory filters or channels. These auditory channels are thought to contribute more or less independently to the auditory processing of sound, and have a width between one-third and one-fourth octave. In terms of frequency, this means that these channels along the basilar membrane become wider towards higher frequencies.

While auditory spectral resolution can be classified as rather coarse, auditory temporal resolution is more finely tuned (see Moore, 2003). The auditory system can detect time differences as small as several milliseconds (see e.g. Davis and McCroskey, 1980), which indicates that auditory processing is many times faster than, for instance, the human eye. Temporal resolution is mainly determined by the response time of the basilar membrane within a specific auditory band, but is also enhanced by across-channel analysis of the sound.

3. Speech as we hear it

It may be argued that the real understanding of speech only begins in the central auditory and non-auditory nervous systems in the human brain. At this level, the speech information, as coded in the nerve potentials, is transferred to the perception of phonemes, words, or concepts. However, there does not appear to be a one-to-one relation between the acoustical, spectro-temporal properties of the presented speech and for instance the perceived phonemes. Instead, perception appears to depend on word context (Hillenbrand et al., 2001), speaking condition and speaker identity. Nevertheless, vowels and consonants that are produced by different speakers under different conditions are still perceived as constant categories. This is one of the reasons that phonemes are often considered the smallest unit present within a speech signal (Liberman et al., 1957).

Speech perception at this level is affected by the listener's knowledge of the lexical, grammatical and syntactical constraints of a specific language. When a listener is familiar with the language, i.e. has a high language proficiency (Van Wijngaarden, 2004), speech perception will generally be easier and less information will be needed to still be able to understand what is said. Moreover, speech reception is affected by an individual's working memory capacity and speed of information processing (see, for instance, Hällgren, 2005). Thus, speech intelligibility performance depends on an interaction between, on the one hand, bottom-up or 'stimulus-driven' processes, and, on the other hand, top-down or 'knowledge-driven' factors (Goldstein, 2002). The relative contribution of auditory and non-auditory processes to speech reception is, however, still under discussion. Results described in this thesis contribute to this discussion.

C. Disturbed speech for normal-hearing listeners

Everyday listening conditions typically include combinations of non-stationary competing noise, reverberation and limited bandwidths, for instance by telephones. All these factors may adversely affect speech intelligiblity. They will thus have to be taken into account when trying to predict speech intelligibility, which can be useful, for example, in the design of public address systems, conference rooms or public facilities in general. Examples of speech intelligibility models are the Speech Transmission Index (STI) or the Speech Intelligibility Index (SII).

The Speech Transmission Index (STI; Houtgast et al., 1980) is based on the observation that speech intelligibility is related to the preservation of the envelope fluctuations of speech. This preservation is characterized by the Modulation Transfer Function (MTF), which quantifies the detrimental effects of distortions on the modulations of the envelope of (band-filtered) speech. The STI does not take individual properties of talkers or listeners into account, but is purely a measure of the transmission channel or, to be more specific, the acoustic characteristics of the environment. It has been shown to be strongly related to speech intelligibility and is widely used as an acceptability criterion in room acoustics or telecommunication.

The Speech Intelligibility Index (SII; ANSI, 1997) is in fact a very similar model, but its applicability is essentially restricted to assessing the effect of noise. Instead of the modulation transfer functions, the spectra of speech and noise are used to determine the effective signal-to-noise ratio in each spectral band, after which wheighing over spectral bands gives the SII. It is often interpreted as the amount of undistorted speech information available for an individual listener. The listener's audiogram may be included in the calculation, and different spectral weightings can be chosen to assess various talker styles or speech materials. Moreover, the Extended Speech Intelligibility Index (ESII; Rhebergen and Versfeld, 2005) was recently introduced, which makes it possible to apply the SII in non-stationary backgrounds by calculating and averaging the SIIs determined in short time frames.

The interpretation of the obtained indices is the same for the SII and the STI: poor intelligibility is associated with values below 0.45 and good intelligibility assumes an index larger than 0.60. In the assessment of speech intelligibility for individual listeners, the Speech Reception Threshold (SRT; Plomp and Mimpen, 1979) is commonly used, defined as the signal-to-noise ratio that the listener needs to correctly repeat 50% of the presented sentences correctly. For normal-hearing listeners, the SRT corresponds to a STI or SII of about 0.33. This value may vary due to differences in speaker styles (Van Wijngaarden and Houtgast, 2004). In situations with stationary noise without reverberation, both the SII and the STI predict a SRT of about -5 to -4 dB SNR, consistent with results as found in literature.

In most listening situations, however, the background noise is non-stationary in character: the level of the background noise fluctuates over time. In these conditions, normal-hearing listeners are able to make use of the relatively silent periods or gaps to improve their speech intelligibility performance. The obtained SRT in non-stationary maskers is therefore often substantially lower (better) than in stationary maskers, reaching values from -12 up to -20 dB SNR, depending on the characteristics of the background noise. The difference between the SRT in non-stationary and stationary noise is often referred to as the release from masking due to masker fluctuations, or simply 'masking release'.

The models of speech reception discussed above are unable to deal with the non-stationary character of everyday background noises, in combination with reverberation. Part of this thesis is therefore dedicated to the introduction of a model to make predictions for speech intelligibility in such daily listening situations.

D. Disturbed speech for hearing-impaired listeners

A common complaint of hearing-impaired listeners concerning speech intelligibility is that they 'can hear *that* something is said, but not *what* is said'. Even when the speech is presented at levels well above the hearing threshold, it is still hard for many hearing-impaired listeners to understand the sentences, especially in conditions that involve reverberation, noise or the combination of both.

1. Noise and hearing loss

Specifically in non-stationary background noise, the difference in speech intelligibility performance between normal-hearing and hearing-impaired listeners is remarkable. Hearing-impaired listeners appear to obtain less benefit, or no benefit at all, from using information present in the gaps. In most cases, merely a lack of audibility, due to audiometric hearing loss or masking noise, is not enough to explain the problems experienced when listening to speech in fluctuating noise (see e.g. Eisenberg et al., 1995).

Plomp (1978) formulated a model description for the speech reception threshold (SRT) based on two auditory parameters: i) hearing loss due to attenuation, related to a raised hearing threshold, and ii) hearing loss due to distortion, which is considered to reflect supra-threshold deficits in hearing (Stephens, 1976). Examples of such supra-threshold deficits are reduced temporal and spectral auditory resolution.

Both reduced temporal resolution and reduced spectral resolution are thought to adversely affect speech perception in noise, especially in fluctuating maskers (Festen and Plomp, 1990; Dubno et al., 2003). Some studies, however, indicate that even listeners with significantly broadened spectral filters still have sufficient spectral resolution to resolve the spectral cues important for speech intelligibility (Ter Keurs et al., 1993a,b).

Besides audibility and supra-threshold deficits, a third factor that may adversely affect speech reception for hearing-impaired listeners is the ability to reconstruct the semantic structure of speech from incomplete information. As demonstrated by Warren (1970), young normal-hearing listeners can perceive missing phonemes by using the redundancies in speech at the acoustic, phonetic, phonological and / or lexical level. This ability is considered to be of great importance when the speech signal is partly masked or disturbed. Grant et al. (1998) referred to this function as perceptual closure, i.e. the ability to form linguistic wholes from perceived fragments. This function, and other cognitive or verbal processing functions relevant for speech reception like working memory capacity or processing speed, may well be reduced for hearing-impaired listeners, who, in addition to their hearing loss, are also often older. CHABA (1988) presents a review of the effects of aging on the processes relevant for speech reception, and the resulting methodological difficulties due to a confounding of the effects of age and hearing loss.

2. *Reverberation and hearing loss*

Even though reverberation is as common and as troublesome for hearing-impaired listeners, most research has focused on the adverse effects of background noise (CHABA, 1988). One of the reasons for this attentional focus may be that adding noise to a speech signal is a simpler and more practical operation than introducing reverberation. Moreover, the adverse effects of reverberation on speech intelligibility may often be quantified by applying the Speech Transmission Index (STI). The influence of reverberation can then be translated into an equivalent speech-to-noise ratio, after which it is treated in essentially the same way as additive noise.

This approach has been shown to be valid for normal-hearing listeners: a certain reduction in STI always gives rise to a corresponding reduction in sentence intelligibility (Houtgast and Steeneken, 1984), whether the speech degrading factor is reverberation, noise, spectral filtering or something else. Measurements by Duquesnoy and Plomp (1980) confirmed that the STI can also be applied for a group of listeners suffering from presbycusis. However, the STI-method does not necessarily apply to other groups of hearing-impaired listeners. Individual listeners may vary widely in their susceptibility to noise and reverberation and the mechanisms underlying the intelligibility of speech in noise or in reverberation may even be different (Nabelek and Dagenais, 1986).

Moreover, assessing the effect of reverberation on speech reception by measuring the SRT in noise at various reverberation times (like Duquesnoy and Plomp, 1980) always gives rise, by definition, to a confounding effect of the presence of noise. This thesis introduces a test that enables an efficient examination of the sole effect of reverberation on speech intelligibility.

E. Outline of this thesis

It may be clear from the discussion above that speech reception in everyday listening situations depends on both the auditory and non-auditory processing abilities of the listener (internal factors), and on the acoustic characteristics of the environment (external factors). This thesis basically deals with both these components. More specifically, the aims of the research presented in this thesis are:

i. to investigate why hearing-impaired listeners benefit less than normal-hearing listeners from the relatively silent periods or 'gaps' in fluctuating noise; that is, why they obtain a smaller release of masking when the masker level fluctuates.

ii. to investigate whether the relative contributions of noise and reverberation to disturbing speech intelligibility depend on the type of hearing-impairment.

In the first part of the thesis, the effect of non-stationary noise on speech intelligibility is investigated. **Chapter 2** discusses an experiment in which Speech Reception Thresholds (SRTs) in stationary noise and in several amplitude-modulated noises were measured for normal-hearing listeners, sensorineural hearing-impaired listeners, and normal-hearing listeners with simulated hearing loss. The latter group consisted of normal-hearing listeners, who received an additional masking noise, giving rise to an artificially raised hearing threshold. This approach made it possible to determine whether differences in masking release are due to a loss of signal audibility, or due to supra-threshold deficits, such as reduced spectral and temporal resolution. Results show that the reduced masking release can only partly be accounted for by reduced audibility. Instead, temporal resolution and age are shown to be the main factors governing masking release for speech in modulated noise, accounting for more than half of the inter-subject variance.

It is suggested that the observed adverse effect of age on the advantage in intelligibility obtained from 'listening in the gaps' may be associated with cognitive or other non-auditory factors. **Chapter 3** investigates this suggestion, discussing an experiment in which auditory measurements were performed for normal-hearing and hearing-impairment listeners. In addition, a visual analogue of the SRT test was included in this experiment: the Text Reception Threshold test or TRT. Results show that nonauditory factors are indeed important for the comprehension of sentences in stationary and non-stationary background noise.

Chapter 4 extends these results with measurements in a larger group of age-matched normal-hearing and hearing-impaired participants, also including tests of cognitive non-auditory functions. Results show that the obtained age-effect is indeed related to the importance of non-auditory processing for speech reception in fluctuating noise. In particular, a test of spatial working memory (SWM) and the TRT appeared to be relevant for the comprehension of speech in fluctuating backgrounds.

It will be shown that this finding makes it possible to investigate the sources responsible for deteriorated speech reception in clinical practice.

In the second part of the thesis, the combined effects of noise and reverberation are considered. In **Chapter 5**, a model is introduced to predict SRTs for normal-hearing listeners in non-stationary noise and reverberation, making use of the Extended Speech Intelligibility Index (E-SII) and the Speech Transmission Index (STI). After taking the characteristics of the speech corpus into account, results show that the model accurately predicts SRTs for combinations of fluctuating noise and reverberation for normal-hearing listeners.

Chapter 6 discusses the development of a simple adaptive test, the Speech Reception Reverberation Threshold, designed to determine the amount of reverberation that an individual listener can sustain to still understand 50% of the presented sentences. This test is validated and its application for assessing hearing-impairment is discussed.

Chapter 7 presents the general conclusions from this thesis and discusses the clinical relevance of the current findings. Moreover, some suggestions are given for future research to further investigate the processes underlying differences in speech intelligibility between normal-hearing and hearing-impaired listeners.

It should be noted that this thesis is composed of five papers (Chapters 2 to 6), published or (to be) submitted for publication as a research paper. This means that these chapters can be read separately, but also has as a consequence that there may be some overlap in the Introduction, Methods or Results sections of these chapters.

Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners

2

Erwin L.J. George, Joost M. Festen and Tammo Houtgast

Journal of the Acoustical Society of America (2006), **120**, 2295-2311.

The Speech Reception Threshold for sentences in stationary noise and in several amplitude-modulated noises was measured for 8 normal-hearing listeners, 29 sensorineural hearing-impaired listeners, and 16 normal-hearing listeners with simulated hearing loss. This approach makes it possible to determine whether the reduced benefit from masker modulations, as often observed for hearing-impaired listeners, is due to a loss of signal audibility, or due to supra-threshold deficits, such as reduced spectral and temporal resolution, which were measured in four separate psychophysical tasks. Results show that the reduced masking release can only partly be accounted for by reduced audibility, and that, when considering supra-threshold deficits, the normal effects associated with a raised presentation level should be taken into account. In this perspective, reduced spectral resolution does not appear to qualify as an actual supra-threshold deficit, while reduced temporal resolution does. Temporal resolution and age are shown to be the main factors governing masking release for speech in modulated noise, accounting for more than half of the inter-subject variance. Their influence appears to be related to the processing of mainly the higher stimulus frequencies. Results based on calculations of the Speech Intelligibility Index in modulated noise confirm these conclusions.

I. INTRODUCTION

A common complaint among hearing-impaired listeners is that speech, although audible, may not be understood, especially in conditions where background noise is involved. When masker levels fluctuate over time, the difference between normal and hearing-impaired listeners is remarkable. Hearing-impaired listeners appear to benefit less from the relatively silent periods or gaps in this type of masker (Festen and Plomp, 1990; Bronkhorst and Plomp, 1992; Takahashi and Bacon, 1992; Hygge et al., 1992; Gustafsson and Arlinger, 1993; Stuart and Phillips, 1996; Peters et al., 1998; Snell et al., 2002; Nelson et al., 2003; Wagener and Brand, 2005). Understanding the processes behind this effect is important, since fluctuating backgrounds are very common in everyday situations.

A. Factors influencing masking release

It is still unclear to what extent audiometric differences can account for the above mentioned differences in masking release. In most cases, merely a lack of audibility, due to audiometric hearing loss or masking noise, is not enough to explain the problems experienced when listening to speech in fluctuating noise (Eisenberg et al., 1995; Bacon et al., 1998; Summers and Molis, 2004). As already observed by Plomp (1978), many hearing-impaired listeners have difficulties understanding speech in noise, even if both speech and noise are well above threshold. This motivated a model description for the Speech Reception Threshold (SRT) based on two parameters: hearing loss due to attenuation and hearing loss due to distortion. When speech and noise have similar overall spectra and are above threshold for all frequencies, the distortion component is considered to be the reflection of supra-threshold deficits in hearing. These deficits are considered to be caused by deterioration of the functioning of the inner ear, like reduced temporal and spectral resolution and a loss of normal auditory compression. The inter-relationship between these deficits and their relation with the hearing threshold is still under discussion (Ludvigsen, 1985; Moore et al, 1999; Oxenham and Bacon, 2003).

Reduced temporal resolution is known to adversely affect masking release. A loss of temporal resolution gives rise to more forward masking, i.e. the masker will decay slower after termination of the masking sound, thus decreasing perceived gap size. Therefore, hearing-impaired listeners with reduced temporal resolution are expected to experience a smaller masking release in speech (Glasberg et al., 1987; Festen and Plomp, 1990; Glasberg and Moore, 1992; Festen, 1993; Dubno et al., 2003). In these studies, however, the amount of variance in masking release accounted for by measures of temporal resolution remains unspecified.

The influence of reduced spectral resolution on masking release is less clear, although spectral resolution plays a central role for speech intelligibility in noise (Celmer and Bienvenue, 1987; Healy and Bacon, 2006). Results from two studies by Baer and Moore (1993, 1994) show that loss of spectral resolution is related to reduced masking release. Measurements simulating cochlear implants also confirm the importance of spectral resolution to masking release (Nelson and Jin, 2004; Xu et al., 2005). In contrast, however, ter Keurs et al. (1993a,b) show that reduced spectral resolution in hearing-impaired listeners is only loosely associated with speech intelligibility in noise, although a significant influence of spectral smearing on masking release was found. They conclude that even listeners with significantly broadened filters still have sufficient spectral resolution to resolve spectral cues important for speech intelligibility.

Besides audibility and supra-threshold deficits, a third factor that may affect masking release, mentioned by e.g. Füllgrabe (2006), is a listener's general ability to reconstruct the spectro-temporal structure of speech from incomplete information. As demonstrated by Warren (1970), human listeners can perceive missing phonemes by using the redundancies in speech at the acoustic, phonetic, phonological and / or lexical level. More recently, Howard-Jones and Rosen (1993) showed that normal-hearing listeners are able to integrate information across various spectral bands at different times, a process which they called 'uncomodulated glimpsing'. This ability is likely to be adversely affected in listeners with deteriorated spectral or temporal resolution. Other effects that may affect masking release are comodulation across spectral bands (Hall, 1984) and informational masking (Summers and Molis, 2004). Particularly comodulation across bands (comodulation masking release or CMR) may also be related to deteriorated spectral or temporal resolution (Hall et al., 1988), although it is expected to only slightly influence masking release (Festen, 1993).

B. Accounting for audibility

A common approach to distinguish between problems in speech reception related to limited audibility and to supra-threshold deficits is to include not only normal-hearing (NHR) and hearing-impaired (HI) listeners in an experiment, but also a third group of subjects with only threshold-related problems. In the current experiment, this group consisted of normal-hearing listeners who received an additional masking noise, such that their masked pure-tone thresholds were equal to the average hearing threshold in the hearing-impaired group at 1/3-octave frequencies (cf. Fabry and Van Tassell, 1986; Humes et al. 1987; Zurek and Delhorne, 1987; Dubno and Schaefer, 1992). This group with simulated hearing loss will be indicated as the SIM-group.

When listening to speech, hearing-impaired listeners may suffer from reduced audibility and supra-threshold problems. However, listeners with a simulated hearing loss, by definition, only suffer from a threshold-related problem. Therefore, comparison of the differences in the SRT between these groups makes it possible to distinguish between threshold-related and supra-threshold problems in understanding speech. Supra-threshold problems in speech understanding are defined here as the specific part of individual deterioration in speech intelligibility performance, that cannot be accounted for by a loss of audibility. The difference in masking release between the NHR-group and the SIM-group is then regarded as an estimate for the threshold-related component of the speech hearing loss, while the difference between the HI-group and the SIM-group is considered to be an estimate of the component due to supra-threshold deficits.

This approach was pursued earlier by Bacon et al. (1998), who measured masking release in temporally complex backgrounds for normal-hearing, hearing-impaired and noise-masked normal-hearing listeners. They concluded that reduced masking release in speech for hearing-impaired listeners can only sometimes be accounted for entirely by reduced audibility. Similar conclusions can be drawn from other studies (Zurek and Delhorne, 1987; Dubno and Schaefer, 1992; Eisenberg et al., 1995; Dubno et al., 2002, 2003), although the contribution of audibility and specific supra-threshold deficits to speech understanding still remains unclear.

An alternative method to account for the effects of audibility is to spectrally adapt the auditory stimuli, to assure audibility of the signal at all frequencies. This method was also applied in the current study: all measurements were performed in two spectral modes, with the masker and the speech spectrally shaped according to the long-term average of natural speech, or optimized with respect to individual hearing thresholds. Differences in results between the two modes will be investigated to assess the influence of spectrum shape on speech recognition.

A drawback of spectrally adapting the signal with respect to individual hearing thresholds is, however, that the overall level of the signal is different for each listener. Effects of presentation level will thus have to be considered. It is known for normal-hearing listeners that masking release, spectral resolution and temporal resolution are level-dependent. Spectral resolution deteriorates with increasing presentation level for normal-hearing listeners (Dubno and Schaefer, 1992; Sommers and Humes, 1993a,b), while temporal resolution is enhanced at higher levels (Jesteadt et al., 1982; Fitzgibbons, 1983; Fitzgibbons and Gordon-Salant, 1987). Masking release is expected to be positively affected by the better temporal resolution at higher levels (Festen, 1993). However, Summers and Molis (2004) showed some evidence that benefit of masker fluctuations decreased for levels above 60 dB SPL. To adequately distinguish between threshold-related and supra-threshold deficits, the effect of level on masking release and on spectral and temporal resolution as found in the current experiment will be taken into account before investigating the influence of the actual supra-threshold deficits on masking release.

Finally, the effect of audibility can be accounted for by applying the Speech Intelligibility Index or SII (ANSI S3.5-1997), a measure of speech intelligibility performance which is able to handle inter-subject audiogram and spectrum differences. The SII has been extensively validated for stationary masking noise and recently, Rhebergen & Versfeld (2005) proposed an extension to the model, that makes it also applicable to fluctuating background maskers. A slightly modified version of their model will be used in the current study to translate the measured SRT-values to SII-values. The SII will be introduced further in the Section D of the Results. Validation and details of the SII model can be found in the Appendix.

C. Objectives

The objectives of this paper are (1) to determine the magnitude of the differences between normal-hearing and hearing-impaired listeners in masking release for speech across various conditions of modulated noise, and (2) to investigate the extent to which differences in masking release between groups can be accounted for by reduced audibility and by supra-threshold deficits. Thus, although results in the various conditions will be reported, this paper will focus on differences in masking release between the groups, instead of differences between conditions. Therefore, the results will be collapsed across the various modulation characteristics, to obtain an overall measure of masking release for speech due to masker modulations.[*]

[*] Prior to collapsing, the individual conditions were investigated by performing a repeated-measures ANOVA on the data from all hearing-impaired participants. This analysis showed that the effects of none of the independent variables on masking release change over the various conditions, indicating that collapsing over conditions when investigating these effects may be considered reasonable.

It is expected, consistent with the literature mentioned above, that differences in hearing thresholds between groups will not be enough to explain differences in masking release. Therefore, correlations will be studied between differences in masking release and four measures of spectral and temporal acuity, as measured for all individual listeners. Two of these measures (spectral resolution (F) and temporal resolution (T)) are derived from psychophysical detection tasks and are regarded to estimate the individual auditory-filter and temporal-window width around the central frequency of 1 kHz. The other two (Speech Reception *Bandwidth* Threshold (SR*B*T) and Speech Reception *Timewidth* Threshold (SR*T*T)) use a procedure similar to the SRT-test to estimate the amount of speech information needed in a limited frequency range and in short time intervals, respectively. The reason for measuring the SR*B*T and the SR*T*T is that they reflect a listener's ability to extract speech information from a spectral or a temporal gap in noise (Noordhoek et al., 1999, 2000). They are considered to be related to spectral and temporal resolution (F and T), but also to the earlier mentioned ability to reconstruct the spectro-temporal structure of speech from incomplete information by using redundancies at the acoustic, phonetic, phonological and / or lexical level (cf. Warren, 1970). As such, they may involve capacities that are relevant for masking release in modulated noise.

II. EXPERIMENT AND METHOD

A. Apparatus

The experiment was run on a Dell personal computer, with a Creative Labs Audigy external sound device and Beyer Dynamic DT48 headphones. To be able to reach high stimulus levels, an additional Shure FP22 stereo headphone amplifier was used. All measurements were performed while listener and investigator were seated in a sound-insulated room. Interfering signals were generated by multiplying white noise with the appropriate amplitude modulation function, after which the speech spectrum was imposed by filtering with a 2048-point finite impulse response (FIR) filter. Final spectral shaping of both speech and masker was performed via a 1024-point windowed FIR filter, by using individual thresholds as inputs. This filter also corrected the frequency response of the headphones and restricted the bandwidth of both noise and speech signal to frequencies between 125 Hz and 8 kHz.

B. Speech material

A set of short meaningful everyday sentences was used, as developed and evaluated by Versfeld et al. (2000). The first 32 lists of this set, read by a male speaker, were

nr.	masker description	duty-cycle (dc) [%]	mod. depth (md) [dB]	mod. freq. f_{MOD} [Hz]
1	Silence	-	-	-
2	Stationary	100	0	0
3	Speechmodulation	undefined	speechlike	speechlike
4	Block, default	50	∞	16
5	Block, dc = 75%	***75***	∞	16
6	Block, md = 15 dB	50	***15***	16
7	Block, f_{MOD} = 32 Hz	50	∞	***32***
8	SR*T*T (fixed SNR)	variable	∞	16

Table 2-I Details on the temporal characteristics of the masking noises. For the block-modulated maskers, non-default values are displayed in bold-italics. The temporal waveforms of the various background maskers are shown in Figure 2-1. Not mentioned in this table is the SRBT, which was measured for a subset of listeners. For more details on the SRTT and the SRBT, see the text.

used, each list containing 13 sentences of eight or nine syllables. The set was developed to enable efficient measurement of the SRT in stationary speech-shaped noise and can be considered as being equivalent to the smaller sentence set of Plomp and Mimpen (1979), giving rise to a standard error in SRT of about 1 dB for normal-hearing listeners. Under all masker conditions, the long-term spectra of the speech and the masker were similar in shape.

C. Interfering signals

SRT-measurements were performed using a variety of background maskers, differing both in temporal and spectral characteristics. Eight different temporal masking conditions were applied in two modes, with the masker and the speech spectrally shaped according to the long-term average of natural speech, or optimized with respect to individual hearing thresholds. The long-term spectral shape of the masker was equal to that of the speech, in all conditions. All tests were conducted both in test and retest, so, a total of thirty-two (8 backgrounds * 2 spectral modes * 2 tests) SRT-measurements were performed for each participant. In addition to the set of thirty-two SRT-measurements, the SR*B*T (Noordhoek et al., 1999, 2000) was measured for a subset of listeners.

1. Temporal characteristics

SRT-measurements were performed in background noise with temporal characteristics ranging from stationary noise to fast block-wave modulated noise. Masker characteristics were chosen to investigate the different aspects of relatively silent periods to speech intelligibility as adequately as possible. Detailed information about the used conditions is given in Table 2-I. The temporal waveforms of the various background maskers are shown in Figure 2-1.

The silence (1) and stationary-masker (2) conditions were included as reference conditions. In condition (3), a masker was used that specifically mimics the intensity fluctuations of speech. This masker was generated using the method described by Festen & Plomp (1990), which splits up a steady-state masker in a low- and high-frequency part with 1000-Hz crossover frequency. Both parts are then modulated separately with the envelope of speech from the corresponding frequency region,

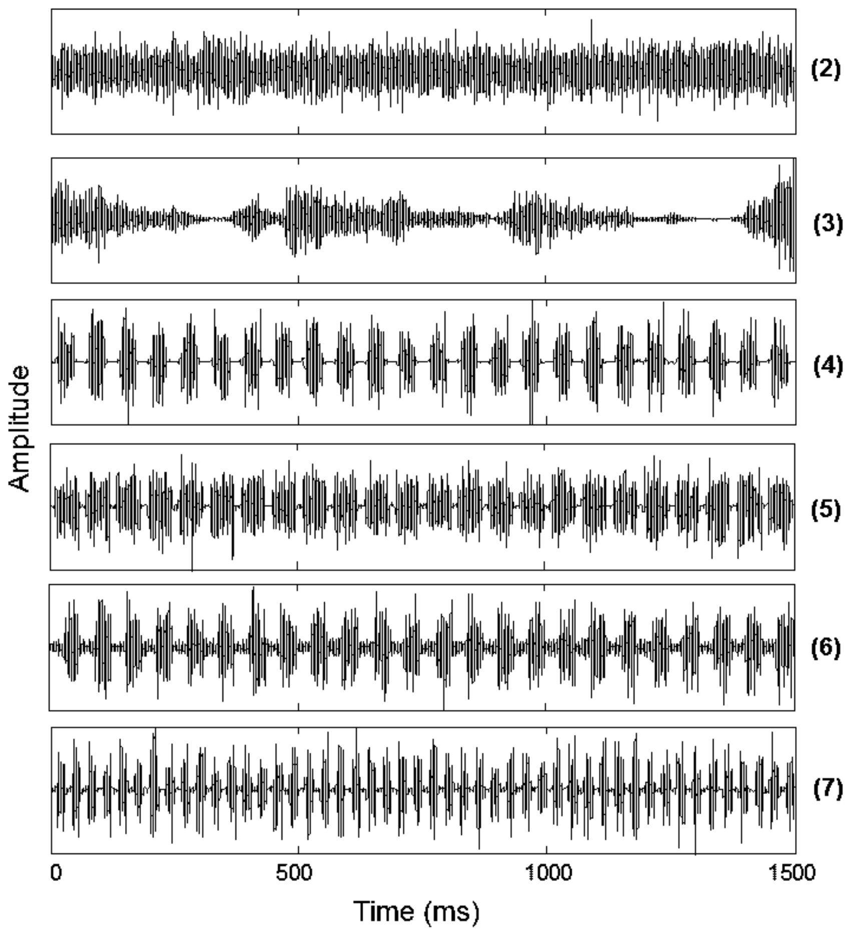

Figure 2-1 The temporal waveforms of the various background maskers: stationary noise (2), speech-modulated noise (3) and four different forms of block-modulated noise (4 to 7). The temporal waveform of the SR*T*T-condition (8, not shown) is similar to the waveform of condition (4), only with an adaptively changing duty-cycle.

after which they are added while restoring the original level-ratio between the two bands.

Conditions (4) to (7) use block-modulated noise. To investigate the masking effects of modulated noise, condition (4) was chosen as the default condition, whereas in each of the conditions (5) to (7) one masker-parameter is varied, respectively the duty-cycle (dc), the modulation depth (md) and the modulation frequency (f_{MOD}) of the masker. In condition (4), the masker contains a fairly long silent period, which may possibly lead to a better SRT, when the listener uses the gaps in the noise optimally. The conditions (5) to (7) decrease, each in its own way, the available amount of speech information in the gap, giving rise to a deterioration of the SRT. This approach enables a comparison of SRT's between these conditions to determine the effect of different masker-parameters.

Finally, condition (8) was included to determine the time window width of clear speech, that a listener needs to correctly reproduce 50% of the sentences. In this condition, the masker and the speech are of equal level, and block modulated (chopped) in such a way that they alternate. This means that speech is only present when the masker is absent and vice versa. The difference with a standard SRT-measurement is that the duty-cycle of both masker and speech is varied adaptively, rather than the signal-to-noise-ratio. The speech duty-cycle at which 50% of the sentences is reproduced correctly will be referred to as the SR*T*T.

Unmentioned in Table 2-I is the spectral equivalent of the SR*T*T, the SR*B*T (Noordhoek et al., 1999, 2000), which was measured for a subset of listeners. It is not listed in Table 2-I since it was not included in the original design of the experiment, but was added considering its similarity to the SR*T*T. Like the SR*T*T, it measures the listener's ability to reconstruct speech from fragments, but now in the spectral domain. The adaptive measurement procedure used is the same as in the SR*T*T, but the available speech bandwidth is varied instead of the speech duty-cycle. The SR*B*T is defined as the speech bandwidth at which 50% of the sentences can be reproduced correctly. Speech and noise were presented at the half-way point in the listener's dynamic range, similar to the SRT in adapted spectral mode.

2. Spectral characteristics

All temporal masker-conditions were presented in two modes, either with a spectrum shaped according to the long-term average of natural speech or with a spectrum that was optimized using the individual's absolute thresholds. These two spectral measurement modes will be referred to as SRT-normal (SRT-n) and SRT-adapted (SRT-a), respectively. The long-term spectral shape of the speech was always equal to that of the masker.

In the SRT-n mode, the masker level was set such that the overall level between 125 Hz and 1 kHz was equal to that level in the SRT-a mode, giving rise to only small level differences between the two modes at low frequencies.

In the SRT-a mode, individual hearing thresholds were used to adapt the spectrum of the masker to reach 1/3-octave masker levels equal to the estimated middle of the dynamic range for each listener. The lower limit of the dynamic range was chosen to be the individual pure-tone threshold at each 1/3-octave, while the upper limit was the uncomfortable loudness level (UCL), here chosen at 110 dB SPL for all listeners. The masker level was the mean of these two. Since pure-tone

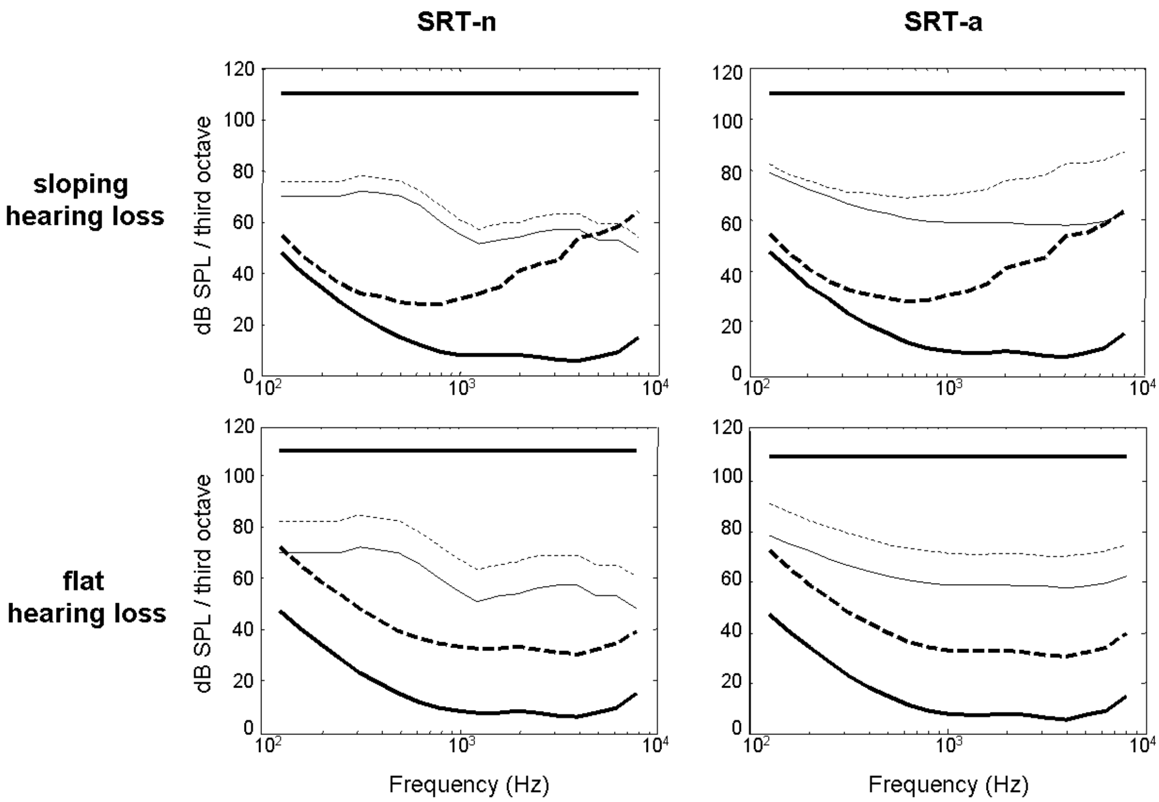

Figure 2-2 Hearing threshold, uncomfortable loudness level (UCL), and the masker spectrum in both normal (left panel) and adapted (right panel) spectral mode. *Solid* lines show the pure-tone hearing threshold (solid bold) and the masker spectrum (solid normal) for a listener without hearing loss. *Dashed* lines show the pure-tone hearing threshold (dashed bold) and the masker spectrum (dashed normal) for a typical listener with sloping hearing loss (upper panel) and flat hearing loss (lower panel). The UCL was set at 110 dB SPL at all frequencies for all listeners and is marked by a straight solid bold line.

thresholds were only measured at 1/3-octave frequencies from 125 Hz to 8 kHz, intermediate threshold levels were calculated by interpolation. The speech signal was shaped accordingly and was varied in level to set the signal-to-noise ratio in the adaptive measurement-procedure.

Figure 2-2 gives an overview of masker spectra in the SRT-n and SRT-a modes, for a normal-hearing listener and two imaginary hearing-impaired listeners with either a typical flat or a typical sloping hearing loss. From this figure, it can be seen that differences between the two spectral modes mainly occur at higher frequencies (above 1 kHz). The SRT-a masker spectrum is above threshold at all frequencies, thus assuring audibility for both normal-hearing and hearing-impaired listeners. This means that in this mode the possible effect of the individual threshold on speech intelligibility is minimized. The SRT-n masker spectrum, on the other hand, approaches the hearing threshold as frequency increases, giving rise to a possible loss of available speech information, mainly in the higher frequencies. This effect occurs especially for listeners with a sloping hearing threshold.

Both spectral modes have their advantages and their drawbacks. Measurement results from the SRT-n mode are bound to give a good impression of the speech intelligibility in real life, since natural spectra are applied. However, inter-individual differences in audibility might influence intelligibility in this mode, making it less powerful in determining effects of supra-threshold deficits. The SRT-a mode minimizes the differences in audibility, and will therefore be more powerful to examine supra-threshold deficits. However, it has the disadvantage that listeners may not be used to listening to sounds in the middle of their dynamic range. In particular, the presence of high-frequency boosted sounds might introduce an unwanted additional disadvantage in understanding speech, especially for listeners with a sloping hearing loss.

D. Procedure

Each experimental session started with the measurement of the listener's pure-tone hearing threshold at all nineteen 1/3-octave frequencies between 125 and 8000 Hz, using the same apparatus used during all other measurements. The audiogram was used later as an input to shape the spectrum of the stimuli. The experiment included a total of thirty-two SRT-measurements (including the SR*T*T), plus the measurement of spectral and temporal acuity and the SR*B*T, defined below. Each measurement was performed twice, following a test-retest design, which makes it possible to estimate reliability for all data points. Measurements were conducted monaurally, using the listeners best ear, which was chosen according to his or her audiogram, or, when the hearing loss was symmetrical, personal preference in telephone conversation.

1. SRT

SRT-measurements were performed using a simple adaptive up-down procedure as described by Plomp and Mimpen (1979). In each condition, the appropriate masker and a list of thirteen sentences, unknown to the listener, were presented. The masker level was kept fixed, while the speech level was varied adaptively to estimate the SRT, defined as the estimate of the speech-to-noise-ratio at which 50% of the sentences could be reproduced without error. In each condition, the first sentence was presented at a level below threshold and repeated, at 4-dB higher levels with each repetition, until the listener was able to reproduce it correctly. The remaining twelve sentences were presented only once, following an one-up-one-down adaptive procedure, with a 2-dB step size. An errorless reproduction of the entire sentence was required for a correct response. The SRT was estimated as the average presentation level of sentences 4 to 13.

Sentences were presented to the listener over the headset and to the investigator visually on a computer screen. To avoid the confounding of both measurement order and sentence lists with condition effects, the order of conditions was counterbalanced across subjects according to an eight-by-eight digram-balanced Latin square, while sentence order was kept fixed. This approach ensures that each of the eight possible temporal conditions, and also pairs of succeeding conditions, occurred only once within each subgroup of eight listeners. For the same reason, half of the participants in each subgroup started with SRT-a conditions, while the other half started with SRT-n conditions.

2. SRTT and SRBT

In the SR*T*T-condition, a comparable adaptive procedure was used. However, in the SR*T*T-condition the speech-to-noise-ratio was fixed, such that the RMS-level of signal and masker were equal. The duty cycle of alternating block-chopped speech and noise was varied in a complementary way to estimate the Speech Reception *T*imewidth Threshold or SR*T*T, defined as the speech duty-cycle (i.e. available timewidth) at which 50% of the sentences was reproduced correctly. Speech was only present when the masker was absent and vice versa, so there was no simultaneous masking. Masker presence was necessary, however, to provide comfortable listening, since chopped speech on its own would be perceived as annoying.

Step sizes of the adaptive SR*T*T-procedure were chosen to fit the step sizes of the standard SRT-procedure (i.e. 4 dB for the first sentence and 2 dB for all other sentences). Classical speech intelligibility prediction models, like the Articulation Index as introduced by French and Steinberg (1947) and Kryter (1962), assume that, when presenting speech to a listener, speech information grows linearly to its maximum over a 30 dB range. Thus, changing the speech level with 4 or 2 dB steps corresponds to a change of available information of 13.3 or 6.7%, respectively. Therefore, duty cycle changes of respectively 12% for the first sentence and 6% for all other sentences seemed appropriate. As in all other conditions, presentation of the first sentence started at a duty-cycle below reception threshold and was repeated, with increasing speech duty-cycle, until the listener was able to reproduce it correctly. All other sentences were presented only once. The SR*T*T was estimated as the average speech duty-cycle while presenting sentences 4 to 13.

Just as the SR*T*T gives an estimate of the amount of speech information needed in short time intervals, the Speech Reception *B*andwidth Threshold or SR*B*T, as introduced by Noordhoek et al. (1999, 2000), gives a good estimate of the amount of speech information a listener needs in a limited frequency range. The SR*B*T-measurement used is the same as Noordhoeks, following the same adaptive procedure as the SR*T*T, but varying the available speech bandwidth instead of the speech duty-cycle. Speech and noise were presented at a level half-way up the listener's dynamic range, similar to the SRT in adapted spectral mode.

Unfortunately, normal-hearing participants were already tested before it was decided to include SR*B*T-measurements in the test battery. This means that they were also not included in the counterbalanced Latin square conditional order. Instead, the SR*B*T was measured after a participant had completed all other measurements.

3. Spectral and temporal resolution

Each listener's spectral and temporal acuities were determined by employing an adaptive measurement procedure as introduced and validated by Hilkhuysen et al. (2005). Validation was performed by measuring eighteen normal-hearing listeners. They showed reliable auditory-filter and time-window widths which were free of noteworthy learning effects, and which varied with presentation level and frequency and corresponded to values as commonly found in the literature. An advantage of the chosen test procedure is that it requires no prior training and can therefore be performed relatively fast (within fifteen minutes) compared to more classical measurement procedures (gap detection, tones in bands of noise). Moreover, the task to perform is relatively simple ("count the number of sweeps") and easy to explain to naïve listeners. In the current experiment, both temporal and spectral acuity were determined in the frequency region around 1 kHz, at a level half-way up the listener's dynamic range (as in adapted spectral mode, to assure audibility).

The measurement procedure included three tests, in which listeners were asked to report the number of tone sweeps (0, 1, 2 or 3) they were able to detect in: i) steady-state noise without grid; ii) noise containing a spectral grid with a 50% duty-cycle on a log-frequency scale; iii) noise containing a temporal grid with a 50% duty-cycle. In all three noises, the tone sweeps to detect were sinusoids with a duration of 200 ms, sweeping upward over a range of 1.6 octaves centered around 1 kHz (0.57 to 1.74 kHz) at a speed of 8 octaves per second. Thus, the sweep reached its center frequency after 100 ms. The masker duration was 2.2 s, and the possible tone sweeps could start at 0.6, 1.0 or 1.4 s after masker onset. The phase of the temporal noise grid varied randomly over trials, while the phase of the spectral noise grid always provided a spectral gap logarithmically centered around 1 kHz. Third-octave levels of the masker were set half-way up the listener's dynamic range, and sweep and masker had equal spectrum shape in the frequency range of the sweep.

The level of the tone sweeps (for the steady-state noise) or the gap width of the noise-grid maskers was varied adaptively in a one-up-one-down 4-AFC-procedure (Levitt, 1971), starting above detection threshold for all listeners. In the two noise grids, the level of the tone sweeps was fixed at a signal-to-noise ratio of -6 dB. In steady-state noise, the initial step size was 4 dB, while, in the grid conditions, the gap width was initially changed with a factor $\sqrt{2}$. After four transitions of an incorrect response following a correct response, the actual test started, using step sizes of 2 dB or a factor $2^{(1/4)}$ respectively. The actual test consisted of a random-order set of 24 stimuli, in which each number of sweeps (zero to three) was presented six times. Responses to the zero-sweep trials had no consequences for the adaptive procedure. The detection threshold was defined by the average level or gap width of the last 18 presentations in which one or more sweeps were present. On the basis of the three outcome measures (threshold level in steady-state noise, width of the spectral grid, width of the temporal grid), auditory-filter and time-window widths were estimated

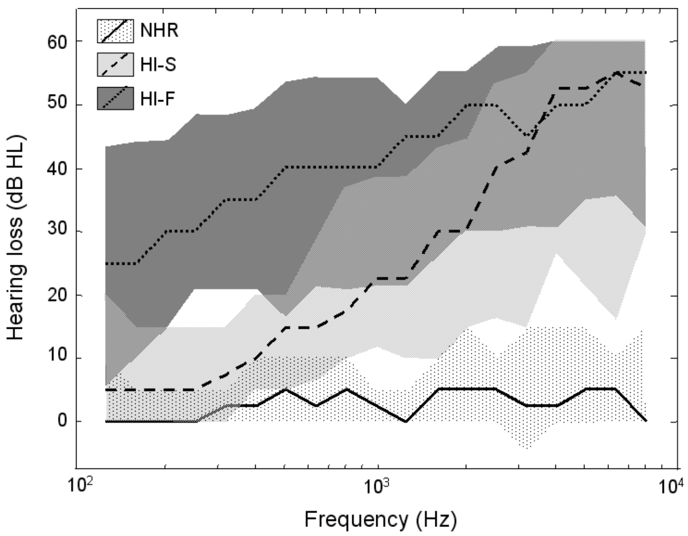

Figure 2-3 Average pure-tone hearing threshold (*re:* ISO-389-1991) and the range between the 5th and 95th percentiles for normal-hearing listeners (NHR) and two groups of hearing-impaired listeners (HI-S and HI-F). The overlap between HI-S and HI-F is displayed as a medium gray region.

by a fitting procedure, assuming a symmetrical one-parameter RoEx-function for the spectral filter shape (Patterson et al., 1982) and a decaying exponential for the time window shape (Duifhuis, 1973; Festen and Plomp, 1981).

E. Participants

1. Normal-hearing listeners (NHR)

The reference group consists of eight normal-hearing listeners, selected to have pure-tone hearing thresholds better than 10 dB HL at .25, .5, and 1 kHz and better than 15 dB at 2 and 4 kHz. Six of the eight participants included in this group were university students. Group age ranged from 19 to 56 years, with an average of 28.5 years.

2. Hearing-impaired listeners (HI)

A group of thirty-two listeners with a sensorineural hearing impairment were selected from the patient database of the audiology department of the VU University Medical Center. Sixteen listeners having a flat hearing threshold (HI-F) and sixteen listeners with a sloping hearing threshold (HI-S) were selected. To be included in the HI-F group, pure-tone hearing thresholds on octave frequencies between .25 and 4 kHz were required to be larger than 25 dB HL, not varying more than 10 dB around their average. The HI-S group was selected to have a pure-tone hearing threshold better than 10 dB HL at .25 kHz, and between 30 dB and 60 dB at 4 kHz. These groups were included to investigate the effect of hearing-loss shape on speech intelligibility.

The final HI-F group included only thirteen of the original sixteen participants identified. One participant did not fulfill the inclusion criteria and two participants in this group showed very large inconsistencies between results on test and retest and were therefore excluded from further analysis. The final thirteen participants in the HI-F group were aged between 27 and 80 years, with an average of 64.5 years. The age of the sixteen participants in the HI-S group ranged from 45 to 76 years, with an average of 60.8 years.

Figure 2-3 displays the hearing threshold for the normal-hearing and for the two groups of hearing-impaired participants, including the regions between the 5th and 95th percentiles. As can be seen, the overlap between both groups of hearing-impaired is fairly large, especially in the higher frequencies. Instead of representing two groups of hearing-impaired with clearly different audiograms, the hearing thresholds seem to be more adequately described as forming a continuum. Therefore, results of the two groups of hearing-impaired will not be dealt with separately, although the distinction is preserved in displaying the results, to enable easy comparison with the two SIM-groups.

3. Simulated hearing loss listeners (SIM)

A group of participants with simulated hearing loss was also included, consisting of sixteen normal-hearing listeners not included in the NHR-group. These listeners' thresholds were elevated by adding noise to the stimuli. The criteria used to select participants in this group were the same as for the normal-hearing reference group.

Similar to the group of hearing-impaired, this group of simulated hearing loss listeners was split in two. In eight listeners, a rather flat hearing threshold was simulated, equal to the average hearing threshold of the thirteen hearing-impaired listeners in the HI-F group. A more sloping hearing threshold, equal to the average hearing threshold of the sixteen hearing-impaired listeners in the HI-S group, was simulated in the other eight listeners. These two subgroups will hereafter be referred to as the SIM-F and the SIM-S group, respectively. The participants in the SIM-F group were aged between 19 and 29 years, with an average of 24.5 years. The age of the participants in the SIM-S group ranged from 19 to 24 years, with an average of 21.8 years.

The hearing threshold simulation was performed by presenting an additional noise to the listeners in the SIM-groups. The goal of this additional noise was to mask the low intensity parts of the signal (cf. Fletcher, 1940; Hawkins and Stevens, 1950), such that the listeners would experience audibility problems when listening to speech. The introduction of an additional broadband noise seems the most appropriate control condition for comparison to sensorineural hearing impairment (Humes et al., 1988). Broadband noise simulates the reduced dynamic range and closely approximates the loudness-growth function of hearing-impaired listeners (Lochner and Burger, 1961; Stevens, 1966). In addition, it has the advantage that performance in the SIM-group can be measured at presentation levels comparable to that of hearing-impaired listeners, although the high level of neuronal activity produced by the noise may not be equivalent to a reduction in activity due to cochlear hearing loss (Fabry and Van Tassell, 1986).

The spectrum level of the additional external noise was chosen to be equal to

$$X = Q - R \qquad (1)$$

where X is the spectrum level of the internal noise representing the elevated threshold, at all 1/3-octave frequencies, Q is the pure-tone threshold level of the audiogram to be simulated (in this case, the average hearing threshold of the appropriate subgroup of hearing-impaired), and R is the critical ratio in dB, as reported by Pavlovic (1987). Since the audiogram measurement at the start of each experimental session was performed in the presence of this background noise, the correct simulation of the desired hearing threshold could be checked. All participants in the SIM-groups showed thresholds within a few dB of the desired threshold, at all nineteen 1/3-octave frequencies.

III. RESULTS AND DISCUSSION

This section gives an overview of the results. In section A, results from SRT-measurements are discussed broadly, to clarify our observation that presentation level is an important factor which should be taken into account when drawing conclusions regarding the effect of supra-threshold deficits on masking release for speech. Section B presents a discussion on the effect of presentation level on tests that are considered to be linked to supra-threshold deficits. These level effects are used to relate the actual supra-threshold deficits to masking release for speech in modulated noise. In section C, a similar exercise is performed for the SRT-values, whereas in section D, the Speech Intelligibility Index (SII) is introduced. The SII is applied to transform SRT to SII values, the results of which are discussed in section E. Finally, section F addresses the difference in results between both spectral modes of signal presentation.

A. Effect of presentation level on masking release (SRT)

The first goal of our work was to determine the differences in SRT between modulated and stationary noise for normal-hearing and hearing-impaired listeners. Figure 2-4 gives an overview of results from the SRT-measurements in both spectral modes. For each participant, SRT_{MOD} is defined as the average SRT over the four block-modulated masker conditions (conditions 4 to 7). As mentioned in the Introduction, the shown results are thus collapsed over the various modulation characteristics, to obtain an overall measure of masking release. SRT_{STAT} is the SRT measured in stationary noise (condition 2). Detailed results are presented in Table 2-II.

The data displayed in Figure 2-4 show rather small spread among listeners for speech intelligibility in stationary noise, while inter-individual differences in SRT are more apparent in modulated noise, as noted earlier by e.g. Bacon et al. (1998) and Versfeld and Dreschler (2002). The difference between SRT_{MOD} and SRT_{STAT} can be regarded as the extent to which an individual listener benefits from the relatively silent periods in modulated noise, i.e. the release from masking. In Figure 2-4, each

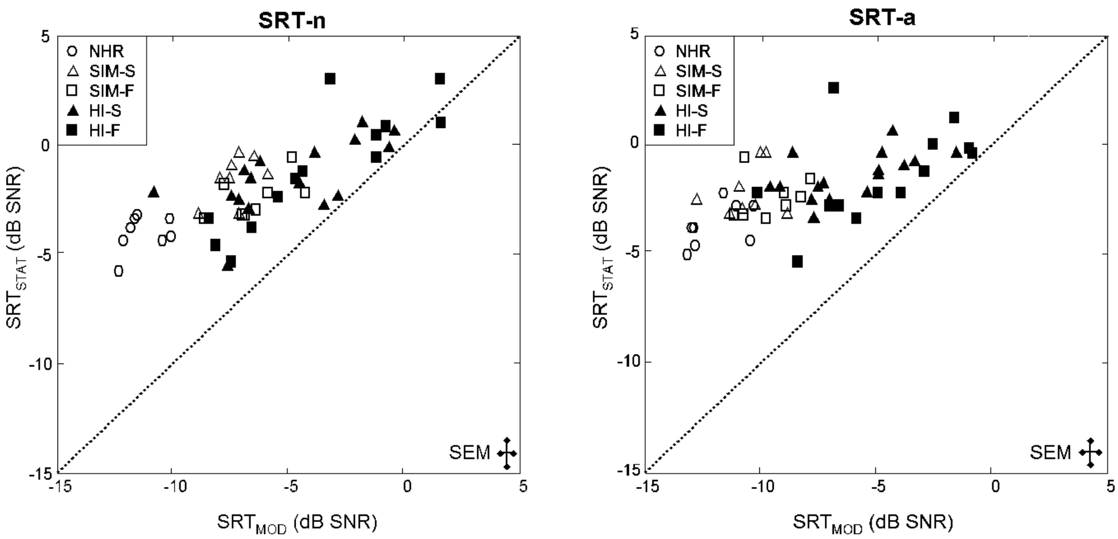

Figure 2-4 SRT in stationary noise versus SRT in modulated noise, defined as the mean SRT from four modulated noise conditions, in both normal (left panel) and adapted (right panel) spectral mode. Data are displayed for normal-hearing listeners (NHR), two groups of hearing-impaired listeners (HI-S and HI-F), and two groups of listeners with simulated hearing loss (SIM-S and SIM-F). The crosses in the lower right corners show the standard errors of measurement (SEM).

individual's vertical deviation from the diagonal is the graphical representation of this benefit. As expected, normal-hearing listeners, denoted by the open circles in Figure 2-4, clearly benefit from the relatively silent periods in modulated maskers, improving their mean SRT from –4.1 dB in stationary noise to –11.3 dB in modulated noise. In contrast, hearing-impaired listeners, as denoted by the closed symbols in Figure 2-4, benefit less when going from stationary to modulated noise; some even obtain no benefit at all.

Listeners with a simulated hearing loss (SIM-S and SIM-F groups, denoted by open triangles and squares) are already distributed closer to the diagonal than normal-hearing listeners, i.e. already have a smaller release from masking than normal-hearing listeners. This indicates that a listener's ability to make use of the relatively silent periods in modulated noise deteriorates, even when no supra-threshold deficits are involved. To put it differently, even when a listener's problems are purely threshold-related, the benefit from the gaps in modulated noise seems to be reduced. This means that a reduced benefit from masker modulations is at least partly linked to threshold-related factors.

To explore this observation further, the difference between SRT_{MOD} and SRT_{STAT} (i.e. the masking release or benefit) is plotted against presentation level in Figure 2-5. More specifically, the negative value of the experienced benefit is plotted along the vertical axis to preserve the analogy to the SRT, so, a more negative value indicates a better performance. Presentation level has been defined here as the long-term A-weighted level of the masker.

To estimate benefit as a function of presentation level, a simple linear regression was performed on data from the normal-hearing and the two SIM-groups, as denoted by all open symbols in Figure 2-5. It was assumed that listeners in these three groups experience no problems related to supra-threshold deficits. Therefore, only the deviation from this regression line represents a reduced masking release possibly due to supra-threshold deficits for a specific listener. This deviation will be referred to as 'Δbenefit', the capital Greek Delta indicating that the effect of presentation level has been accounted for. Data from normal-hearing and simulated hearing loss listeners are distributed close to the regression line, so Δbenefit will be small for these listeners. In contrast, data from most hearing-impaired listeners deviate from the regression line, so their Δbenefit will be larger, indicating a reduced masking release due to supra-threshold deficits.

			normal spectral mode (SRT-n)					adapted spectral mode (SRT-a)				
nr.	masker description	unit	NHR	HI-S	HI-F	SIM-S	SIM-F	NHR	HI-S	HI-F	SIM-S	SIM-F
1	Silence	dB (A)	24.7	39.6	56.0	45.1	63.8	23.0	48.4	57.5	52.4	63.8
2	Stationary	dB SNR	-4.1	-1.6	-1.1	-1.6	-2.5	-3.7	-1.5	-1.5	-2.2	-2.4
3	Speechmodulation	dB SNR	-11.7	-3.3	-2.7	-6.1	-5.5	-11.8	-5.2	-3.5	-9.2	-7.7
4	Block, default	dB SNR	-17.3	-7.8	-5.1	-11.1	-9.1	-17.9	-9.7	-6.9	-15.9	-13.5
5	Block, dc = 75%	dB SNR	-6.4	-2.3	-1.5	-3.9	-4.2	-6.8	-2.9	-2.4	-6.2	-5.8
6	Block, md = 15 dB	dB SNR	-8.8	-4.2	-3.6	-5.6	-5.1	-8.8	-5.2	-4.7	-7.0	-6.8
7	Block, f_{MOD} = 32 Hz	dB SNR	-12.7	-5.6	-4.7	-8.8	-7.5	-14.3	-6.6	-5.2	-13.3	-12.2
8	SR*TT* (fixed SNR)	%	33.3	45.5	48.7	39.0	43.8	31.4	42.5	45.7	34.9	36.0
	SRT_{MOD}	dB SNR	-11.3	-5.0	-3.7	-7.3	-6.5	-11.9	-6.1	-4.8	-10.6	-9.5
	Benefit	dB	-7.2	-3.4	-2.6	-5.7	-4.0	-8.2	-4.6	-3.3	-8.4	-7.1

Table 2-II Details on the temporal characteristics of the masking noises. For the block-modulated maskers, non-default values are displayed in bold-italics. The temporal waveforms of the various background maskers are shown in Figure 2-1. Not mentioned in this table is the SRBT, which was measured for a subset of listeners. For more details on the SRTT and the SRBT, see the text.

It should be noted that the effects of presentation level and hearing threshold cannot be fully distinguished in the current experiment: listeners in the simulated hearing-impaired group were measured at higher presentation levels, but also had an (artificially) raised hearing threshold. So, the regression line in Figure 2-5 shows the combined effects of presentation level and masking noise. In any case, the combined effect needs to be taken into account before considering the effects of supra-threshold deficits on masking release.

In summary, the data show that even listeners with only an audibility-problem (i.e. an artificially raised hearing threshold) experience less benefit than normal-hearing when it comes to understanding speech in modulated noise. Therefore, to study the effects of supra-threshold deficits, it is necessary to eliminate this threshold-related effect.

B. Effect of presentation level on spectral and temporal resolution

The second goal of this study was to investigate the extent to which differences between normal-hearing and hearing-impaired listeners can be accounted for by supra-threshold deficits. Manifestations of individual supra-threshold deficits are the

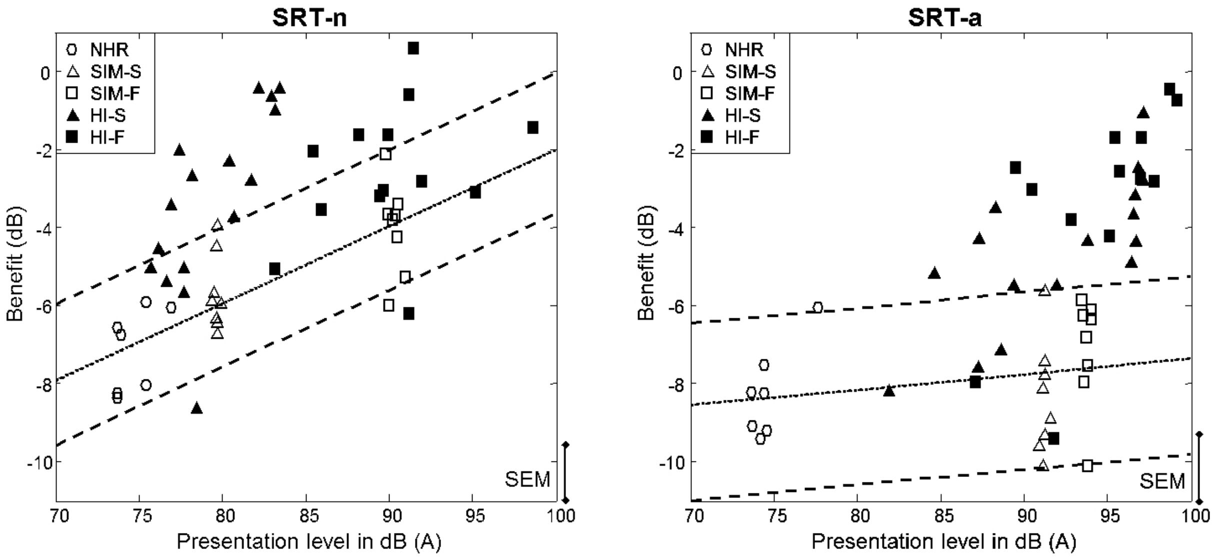

Figure 2-5 Difference between SRT in modulated noise and SRT in stationary noise (i.e. benefit) as a function of masker presentation level, in both normal (left panel) and adapted (right panel) spectral mode. Data are displayed for normal-hearing listeners (NHR), two groups of hearing-impaired listeners (HI-S and HI-F), and two groups of listeners with simulated hearing loss (SIM-S and SIM-F). The dotted lines are regression lines through all open symbols; the dashed lines represents the 5th and 95th percentiles. The bars in the lower right corners show the standard errors of measurement (SEM).

measured spectral resolution (F) and temporal resolution (T), but also the SR*B*T and the SR*T*T (in adapted spectral mode), which are measures of how much speech information is needed to reach 50% intelligibility.

Our second goal can thus be achieved by relating the measured spectral and temporal resolutions, and the SR*T*T and SR*B*T to the experienced unmasking of speech in modulated noise (i.e. Δbenefit). However, as for masking release, the measured resolutions are not independent of presentation level either. This makes it necessary to first investigate these effects. The discussion below will be restricted to spectral and temporal resolution, but also applies to SR*T*T and SR*B*T values. Group averages of the raw data can be found in Table 2-III.

Figure 2-6 shows the measured spectral and temporal resolution as a function of presentation level. Since both spectral and temporal resolution measurements were performed around 1 kHz in the middle of the dynamic range, presentation level has in this case been defined as the masker level in dB SPL per 1/3-octave in the 1-kHz band. Again, a linear regression was performed on data from normal-hearing and simulated hearing loss listeners, who are assumed to experience no problems related to supra-threshold deficits. As before, presentation level is confounded with the presence of masking noise for these listeners. However, since both the tone sweeps

variable	unit	NHR	HI-S	HI-F	SIM-S	SIM-F
F	ERB (Hz)	193.5	234.0	292.7	232.3	291.6
SR*B*T	octaves	(1.4)	1.64	1.87	1.49	1.57
T	ERD (ms)	4.36	4.52	6.68	2.74	2.48
SR*T*T	%	31.4	42.5	45.7	34.9	36.0
age	years	28.5	60.8	64.5	21.8	24.5
PTA	dB HL	4.4	22.8	40.0	23.2	40.9

Table 2-III Group averages of measured spectral resolution F, temporal resolution T, SR*B*T, SR*T*T, age, and pure-tone average (PTA), the latter defined as the average pure-tone hearing threshold at octave frequencies 0.5, 1.0 and 2.0 kHz. Averages were calculated for normal-hearing listeners (NHR), two groups of hearing-impaired (HI-S and HI-F), and two groups of listeners with simulated hearing loss (SIM-S and SIM-F). The average result on SR*B*T for normal-hearing listeners was taken from Noordhoek (1999).

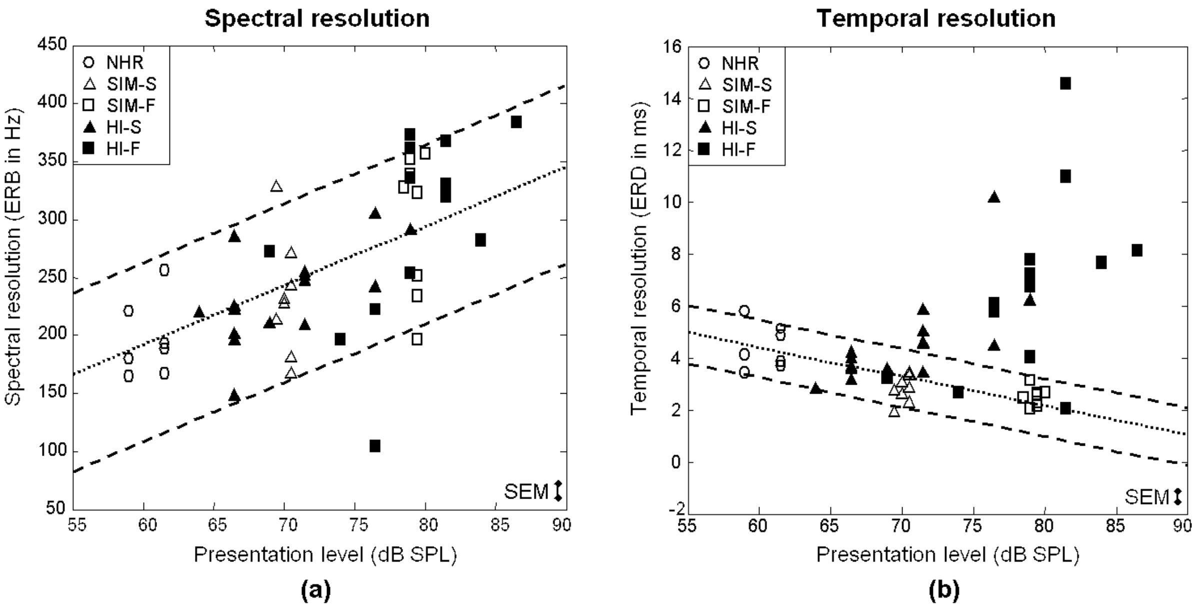

Figure 2-6 Spectral resolution F (panel a) and temporal resolution T (panel b) as a function of presentation level, for normal-hearing listeners (NHR), two groups of hearing-impaired listeners (HI-S and HI-F), and two groups of listeners with simulated hearing loss (SIM-S and SIM-F). The dotted lines are regression lines through all open symbols; the dashed lines represent the 5th and 95th percentiles. The bars in the lower right corners show the standard errors of measurement (SEM).

and the noise grids in the measurements were presented strictly at a level half-way up the dynamic range of a listener, a raised threshold is expected to play only a minor role. Therefore, the observed effect may be considered solely due to increased presentation level.

Figure 2-6 (a) shows that, for normal-hearing and simulated hearing loss listeners, spectral resolution deteriorates with presentation level. Hearing-impaired listeners, denoted by closed symbols, seem to follow that trend: the spectral resolution of the present group of mild to moderate hearing-impaired listeners does not deviate more from the regression line than results from listeners in the SIM-groups. This indicates that deteriorating spectral resolution can be explained from increased presentation levels. Since noise-masked normal-hearing and hearing-impaired listeners show comparable results, spectral resolution does not appear to qualify as a supra-threshold deficit for the present group of hearing-impaired. So, the fact that hearing-impaired listeners show less finely tuned auditory filters does not appear to be a consequence of damage in the inner ear, but the result of increased presentation levels, as suggested earlier by e.g. Dubno and Schaefer (1992) or Sommers and Humes (1993a,b).

On the other hand, figure 2-6 (b) shows that the deteriorating temporal resolution in hearing-impaired listeners cannot be entirely understood in terms of increased presentation level. Results for the normal-hearing and simulated hearing loss listeners (open symbols), with only an audibility problem, indicate that temporal resolution improves with presentation level. However, the temporal resolution of hearing-impaired listeners (closed symbols) largely deviates from this trend, indicating that their reduced temporal resolution does qualify as an actual supra-threshold deficit (cf. Florentine and Buus, 1984).

From these results, it is concluded that the deviation of measured temporal and spectral resolution from the regression line is the most appropriate measure for the actual amount of supra-threshold deficits of a specific listener. These deviations for spectral and temporal resolution will be referred to as respectively 'ΔF' and 'ΔT'. In an analogous way, the effect of presentation level on SR*T*T and SR*B*T can be taken into account. Linear regressions were performed on SR*T*T and SR*B*T-data from normal-hearing and simulated hearing loss listeners. Individual deviations from these regression lines will be referred to by 'ΔSR*T*T' and 'ΔSR*B*T', respectively.

In summary, it can be concluded that the effect of presentation level needs to be taken into account to get a clear view on the influence of supra-threshold deficits on masking release for speech in modulated maskers. For the present group of mild to moderate hearing-impaired listeners, the deterioration of spectral resolution does not qualify as an actual supra-threshold deficit, since it can be accounted for fully by increased presentation level. In contrast, temporal resolution does qualify as an actual supra-threshold deficit.

C. Relations between supra-threshold deficits and masking release (SRT)

In the previous section, it was shown that presentation level has a large effect on possible manifestations of supra-threshold deficits and that this effect should be taken into account before the relation with masking release can be studied. After applying the necessary corrections, it will now be investigated to which extent the inter-subject differences in masking release (Δbenefit) can be accounted for by actual supra-threshold deficits.

Table 2-IV presents correlations between Δbenefit and the four predictor variables ΔF, ΔSR*B*T, ΔT and ΔSR*T*T, both in normal and adapted spectral mode. A log-transformation was necessary to provide for normality of ΔT. Age and pure-tone

	all		**only HI**	
	Δbenefit (normal)	Δbenefit (adapted)	Δbenefit (normal)	Δbenefit (adapted)
ΔF	.09	.10	.05	.00
ΔSR*B*T	*.20*	*.38* (**)	-.02	.22
log(ΔT)	.44 (**)	.76 (**)	.23	.70 (**)
ΔSR*T*T	.50 (**)	.66 (**)	.16	.41 (*)
age	.57 (**)	.81 (**)	.29	.63 (**)
PTA	.19	.46 (**)	.03	.59 (**)

Table 2-IV Product-moment correlations between Δbenefit and ΔF, ΔSR*B*T, ΔT, ΔSR*T*T, age, and pure-tone average (PTA), in both normal and adapted spectral mode, calculated for all participants (left, N=53) and for the hearing-impaired listeners only (right, N=29). Since no normal-hearing SR*B*T-data are available, the ΔSR*B*T-correlations displayed in italics are calculated including only SIM and HI-data (N=45). Correlations indicated with an asterisk are significant. (*) $p < 0.05$ (**) $p < 0.01$

average (PTA) are included as additional predictor variables, the latter being defined as the average pure-tone hearing threshold at octave frequencies 0.5, 1.0 and 2.0 kHz.

In normal spectral mode, the data in Table 2-IV show that temporal resolution and SR*T*T are highly correlated with masking release from modulated noise (or Δbenefit). Age seems to have an effect too, but, since the normal-hearing listeners as a group were considerably younger than the hearing-impaired group, this correlation may be artificial. It is therefore better to consider only data from hearing-impaired listeners, where none of the mentioned correlations are significant any more. This might have been brought about by inter-individual differences in the audibility of parts of the signal in normal spectral mode. Therefore, correlations with data from the adapted spectral mode are expected to give a better estimation of relevant supra-threshold effects, since in the SRT-a, differences in audibility were minimized by amplifying the signal to levels well above threshold for all frequencies.

Table 2-IV shows that in adapted spectral mode, temporal resolution, SR*T*T and age are significantly related to Δbenefit. Spectral resolution is not significantly correlated, and SR*B*T is only significantly correlated when all listeners are included, the correlation disappearing when looking at data from hearing-impaired listeners only. This means that temporal resolution and SR*T*T appear to be the key factors governing speech unmasking in modulated noise, while spectral resolution and SR*B*T have little effect. Furthermore, the correlation with age, even when considering only data from hearing-impaired listeners, indicates that understanding speech in modulated noise deteriorates with age. Finally, PTA appears to be significantly correlated to speech intelligibility in modulated noise.

However, the correlation analysis performed on speech unmasking by modulations, as presented above, is not ideal. It gives a distorted picture, because the chosen predictor variables are cross-correlated, as can be seen in Table 2-V. This may lead to induced correlations between some predictor variables and speech unmasking. PTA, for instance, is highly correlated with temporal acuity ΔT and might therefore only contribute in predicting speech unmasking via ΔT.

To control for these cross-correlations, a stepwise multiple regression analysis was performed on the data from hearing-impaired listeners (N=29). Reported below are the successive contributing predictor variables and the coefficient of determination R^2, corrected for the available degrees of freedom.

The results from the stepwise regression analysis show that in normal spectral mode, as seen earlier, none of the tests contributes significantly to explaining the variance in Δbenefit, possibly due to large inter-individual differences in the audibility of the signal. In adapted spectral mode, temporal acuity is the most significant term ($p < .0001$), accounting for 46% of the variance in Δbenefit, while age is the second most significant ($p = .0003$), explaining 35% of the variance on its own. When both age and log(ΔT) are included in the regression model, together they account for 64% of the variance in Δbenefit, while the other predictor variables do not contribute significantly ($p \geq .08$). Note that both ΔSR*T*T and PTA are not included anymore, indicating that, indeed, their correlation with Δbenefit was mainly induced by cross-correlations with ΔT and age. Finally, it is worth mentioning that the 64% of explained variance may be an underestimate, considering the non-unity test-retest correlations of the predictor variables.

The large contribution of temporal acuity to explaining variance in masking release is in line with our expectations and with literature (Glasberg et al., 1987; Festen and Plomp, 1990; Glasberg and Moore, 1992; Festen, 1993; Dubno et al., 2003). The fact that, apart from hearing threshold, age is a significant factor, is also consistent with earlier findings (e.g. Gustafsson and Arlinger, 1993; Snell et al., 2002). Increasing evidence exists in literature that age can affect the temporal processing of sounds, independently of hearing loss (Snell, 1997; Grose et al., 2001; Dubno et al., 2002; Gordon-Salant and Fitzgibbons, 2004; Gifford and Bacon, 2005). This effect might be related to the influence of cognitive effects on hearing ability, as suggested

	ΔF	ΔSR*B*T	log(ΔT)	ΔSR*T*T	age
ΔF	.81				
ΔSR*B*T	.68 (**)	.85			
log(ΔT)	.24	.41 (*)	.93		
ΔSR*T*T	.35	.54 (**)	.30	.61	
age	.12	.11	.32	.43 (*)	-
PTA	.32	.57 (**)	.81 (**)	.32	.30

Table 2-V Product-moment cross-correlations between ΔF, ΔSR*B*T, ΔT, ΔSR*T*T, age, and pure-tone average (PTA). Bold values on the diagonal are auto-correlations between test and retest values. Only HI-data were included in the calculations (N=29). Correlations indicated with an asterisk are significant. (*) $p < 0.05$ (**) $p < 0.01$

by e.g. Pichora-Fuller et al. (1995), Watson et al. (1996), Gordon-Salant and Fitzgibbons (1997) and Gatehouse et al. (2003), or to other non-peripheral effects like, for instance, the suppression of neuronal envelope locking, as suggested by Las et al. (2005). These effects may also play a role in the remaining unexplained part of the variance. At present, a new experiment is being conducted to further investigate this possibility.

The fact that the SR*B*T and the SR*T*T do not contribute significantly to explaining variance in masking release may indicate that the ability to use the redundancy of speech (at the acoustic, phonetic or lexical level) only plays a minor role when it comes to obtaining benefit from modulated maskers, even though the correlation of SR*T*T with Δbenefit (r = .41) is significant. The small test-retest reliability of the SR*T*T (r = .61) might be another reason why the SR*T*T does not significantly contribute. Moreover, the ability to restore incomplete speech, as measured by the SR*T*T and the SR*B*T, may be cognitive in nature, and may already have been included in the effect of age, see the significant relation between SR*T*T and age (r = .43) in Table 2-V.

The findings above constitute the main result from this experiment: temporal resolution and age are the essential contributors to explain the inter-individual differences in speech unmasking in modulated noise. Combined, they account for 64% of the variance in Δbenefit.

D. From SRT to SII

In the previous section, each listener's benefit in SRT when listening to speech in modulated noise as compared to stationary noise was calculated. After taking the effect of presentation level into account, Δbenefit has been used as a measure for the unmasking of speech in modulated noise. A major drawback of this approach lies in the fact that each listener has a different audiogram, giving rise to inter-subject differences in audibility. In the adapted spectral mode, it was attempted to overcome this drawback by shaping the spectrum of both signal and masker according to the shape of the individual hearing threshold. However, this in turn means that each participant listened to a different spectrum. Using the SRT to quantify the subjects' ability to perceive speech in noise does not take these inter-subject audiogram and spectrum differences into account.

A measure of speech intelligibility performance which is able to handle inter-subject audiogram and spectrum differences is the Speech Intelligibility Index or SII

(ANSI S3.5-1997), which gives an estimate of the amount of speech information available in a certain condition, using the individual's audiogram and the signal and masker spectrum levels as inputs. Since the SII-model already accounts for audiogram- and spectrum differences among listeners, the calculated SII-value needs no correction for presentation level anymore, as will be shown below.

An SII of about 0.30 to 0.35 is commonly considered to be enough to reach 50% speech intelligibility for normal-hearing listeners. Hearing-impaired listeners generally need more speech information, which is supposed to be caused by less efficient handling of the information due to supra-threshold deficits. Therefore, the amount that the SII is raised may be a good measure of the effect of supra-threshold deficits on speech intelligibility.

The SII has been extensively validated for stationary masking noise and recently, Rhebergen & Versfeld (2005) proposed an extension to the model, that makes it also applicable to fluctuating background maskers. Their approach gives a good account for most existing data, producing SII's around 0.35 for normal-hearing listeners. A slightly modified version of their model has been used here to translate the measured SRT-values in SII-values. Validation and details of this model can be found in the Appendix.

E. Relations between supra-threshold deficits and masking release (SII)

Figure 2-7 gives an overview of the calculated SII-values for individual listeners in both spectral modes. SII_{MOD} is the average SII-value over three 16 Hz modulated masker conditions (see the Appendix), while SII_{STAT} is the SII calculated for stationary noise. The figure shows that normal-hearing participants, as expected, display SII-values of around 0.3 in both stationary and modulated noise. Listeners with simulated hearing loss, experiencing only audibility problems due to their artificially elevated threshold, also display an SII around 0.3 in both cases. Although this was to be expected, it forms a strong contrast with the obtained elevated SRT in this group. This demonstrates the power of applying the SII-method in fluctuating noise, which appears to adequately incorporate problems due to only an elevated threshold.

Many hearing-impaired listeners show supra-threshold problems in stationary noise and modulated noise, as their SII's are raised compared to normal-hearing listeners. The difference between SII_{MOD} and SII_{STAT} can then be regarded as the extent to which an individual listener experiences additional disadvantages when listening to speech in modulated noise compared to stationary noise. Since audibility is already incorporated in the SII, the difference in SII between stationary and modulated maskers (i.e. SII_{DIFF}) seems an appropriate measure for the experienced problem in speech unmasking in modulated noise due to supra-threshold deficits.

Next, it will be investigated to which extent these differences in SII can be accounted for by actual supra-threshold deficits. Table 2-VI gives correlations between SII_{DIFF} and the predictor variables ΔF, ΔSR*B*T, ΔT, ΔSR*T*T, age, and pure-tone average (PTA), in both normal and adapted spectral mode. These results are shown to be comparable to the correlations found between Δbenefit and the six predictor variables, as displayed in Table 2-IV. When, as before to prevent artificial correlations, only data from hearing-impaired listeners are considered, temporal resolution, age, and PTA have the largest effects on speech unmasking, being significantly correlated to SII_{DIFF} in adapted spectral mode. In normal spectral mode, only PTA is significantly relevant.

These results can be regarded as a confirmation of our earlier conclusion that temporal resolution and age are the key factors governing speech unmasking in modulated noise. As before, stepwise multiple regression analysis was performed on the data from hearing-impaired listeners to control for the cross-correlations between the predictor variables (see Table 2-V).

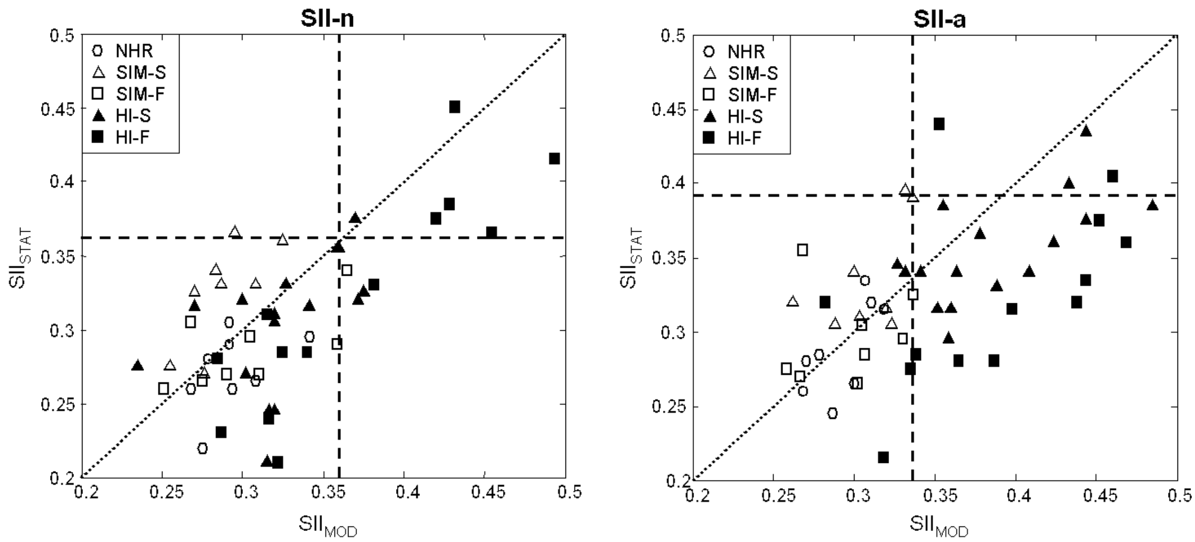

Figure 2-7 SII in stationary noise versus SII in modulated noise, the latter defined as the mean SII in three 16-Hz modulated noise conditions, in both normal (left panel) and adapted (right panel) spectral mode. Data are displayed for normal-hearing listeners (NHR), two groups of hearing-impaired listeners (HI-S and HI-F), and two groups of listeners with simulated hearing loss (SIM-S and SIM-F). Dashed lines represent the one-tailed 95th percentiles for all open symbols.

Results from this analysis show that in normal spectral mode, only PTA significantly contributes (p = .02), accounting for 14% of the variance in SII_{DIFF}. In adapted spectral mode, temporal resolution is the most significant term (p = .0002), accounting for 36% of the variance in SII_{DIFF}, while age is second most significant (p = .0006), accounting for 31% of the variance on its own. When including both predictor variables, they account for 53% of the variance in SII_{DIFF}. As before with SRT, ΔF does not contribute on its own (p = .66), but, rather surprisingly, becomes significant (p = .03) when both temporal resolution and age are included first. When ΔF is included together with ΔT and age, a total of 59% of the variance in SII_{DIFF} can be explained. As in the SRT-case, the other predictor variables do not significantly contribute anymore to the multiple regression ($p \geq .08$).

These findings are consistent with the earlier SRT-results, although the small but significant contribution of spectral resolution is surprising. These SII-results, however, corroborate to our earlier conclusion that temporal resolution and age are the essential contributors to explaining unmasking of speech in modulated background maskers.

F. Difference between normal and adapted spectral mode

In the previous section on SII-results, it is striking that, in normal spectral mode, no predictor variables, except for PTA, have a significant effect on speech unmasking in modulated noise, while they do in adapted spectral mode. In the earlier discussion on the SRT-results, this was explained as a consequence of large inter-individual differences in audibility of part of the signal in normal spectral mode. However, this argument is not valid in the present discussion on the SII-results, since audibility-differences are already accounted for by the SII-calculations.

An alternative explanation for the large difference between SII-n and SII-a results can be found when looking back to Figure 2-7. In normal spectral mode, SII-values for most hearing-impaired listeners are distributed fairly close to normal-hearing data, indicating only small differences between SII_{STAT} and SII_{MOD}. Only six hearing-impaired listeners display a higher-than-normal SII in stationary noise, indicated by a position above the horizontal dashed line. A total of ten hearing-impaired listeners, including these six, display a raised SII in modulated noise. In adapted spectral mode, however, most hearing-impaired SII-values deviate from normal-hearing data. Moreover, SII_{MOD} is larger than SII_{STAT} for most hearing-impaired listeners. Only four hearing-impaired listeners display a raised SII in stationary noise, but no less than 24 out of 29 hearing-impaired listeners display a higher-than-normal SII_{MOD}, indicating that they suffer from supra-threshold deficits when listening to speech in modulated noise.

First, these observations indicate that problems in speech intelligibility due to supra-threshold deficits are more prominent in modulated maskers than in stationary maskers. That is, a listener suffering from supra-threshold deficits will experience more problems understanding speech in modulated maskers than in stationary noise.

Second, these observations indicate that differences in speech intelligibility due to supra-threshold deficits are more prominent in adapted spectral mode than in normal spectral mode. Since differences between stimuli of the two spectral modes mainly occur at the higher frequencies, this finding suggests that mainly the processing of high-frequency stimulus components is affected by supra-threshold deficits, whereas the processing of lower-frequency signals remains relatively unaffected. These findings are consistent with earlier results by Apoux and Bacon (2004), who show that normal-hearing listeners rely more on the higher frequency regions when it comes to the processing of temporal information of speech in noise. Moreover, this result is in agreement with indications by Hogan and Turner (1998), who found that hearing-impaired listeners used the information in higher frequencies less efficiently than normal-hearing listeners, dependent on their degree

	all		**only HI**	
	SII_{DIFF} (normal)	SII_{DIFF} (adapted)	SII_{DIFF} (normal)	SII_{DIFF} (adapted)
ΔF	.09	.07	.00	-.09
ΔSR*B*T	*.19*	*.26*	.08	.15
log(ΔT)	.43 (**)	.67 (**)	.31	.64 (**)
ΔSR*T*T	.26	.50 (**)	-.03	.28
age	.44 (**)	.68 (**)	.25	.60 (**)
PTA	.28 (*)	.37 (**)	.44 (*)	.54 (**)

Table 2-VI Product-moment correlations between SII_{DIFF} and ΔF, ΔSR*B*T, ΔT, ΔSR*T*T, age, and pure-tone average (PTA), in both normal and adapted spectral mode, calculated for all participants (left, N=53) and for hearing-impaired listeners only (right, N=29). Since no normal-hearing SR*B*T data are available, the ΔSR*B*T correlations displayed in italics are calculated including only SIM and HI-data (N=45). Correlations indicated with an asterisk are significant. (*) $p < 0.05$ (**) $p < 0.01$

of hearing loss. Turner and Brus (2001) later suggested that this effect may be linked to the types of speech cues residing in the different regions of the speech spectrum and the consequences of hearing loss upon the tranmission of these cues.

In summary, it can be concluded that inter-subject differences in speech understanding due to supra-threshold deficits are more pronounced in modulated noise than in stationary noise. Moreover, these differences are more prominent in adapted spectral mode than in normal spectral mode, indicating that they are mainly related to the processing of higher stimulus frequencies.

IV. CONCLUSIONS

The main results arrived at in the previous sections can be summarized as follows:

I. Normal-hearing listeners with only an audibility-problem (i.e. an artificially raised hearing threshold) experience a reduced masking release for speech in modulated noise. Therefore, to study the effect of supra-threshold deficits, it is necessary to take the threshold-related effect of presentation level on masking release into account (Section III.A).

II. To get a clear view on the influence of supra-threshold deficits, the effect of presentation level on their manifestations needs to be taken into account. For the present group of mild to moderate hearing-impaired listeners, the observed deteriorated spectral resolution does not qualify as a supra-threshold deficit, since it can be accounted for fully by increased presentation level. In contrast, the observed deteriorated temporal resolution does qualify as an actual supra-threshold deficit (Section III.B).

III. Temporal resolution and age are the essential contributors to explaining the inter-individual differences in masking release for speech in modulated background maskers. Combined, they account for more than half of the inter-subject variance. Results based on SII-calculations confirm this conclusion (Section III.C and III.E).

IV. Applying the SII-method in modulated noise (cf. Rhebergen and Versfeld, 2005) adequately incorporates problems due to an elevated threshold, giving rise to values around 0.3 for normal-hearing listeners with and without an elevated threshold, and larger values for hearing-impaired listeners with supra-threshold deficits (Section III.D / Appendix).

V. Inter-subject differences in speech intelligibility due to supra-threshold deficits are more pronounced in modulated noise than in stationary noise (Section III.F).

VI. Inter-subject differences in speech intelligibility due to supra-threshold deficits are more pronounced in adapted spectral mode than in normal spectral mode, indicating that they are mainly related to the processing of higher stimulus frequencies (Section III.F).

V. ACKNOWLEDGEMENTS

This research was supported by the Heinsius-Houbolt Foundation, The Netherlands. Thanks are due to Gaston Hilkhuysen for providing the tests of temporal and spectral acuities and to two anonymous reviewers for their comments on a previous version of this manuscript.

A. APPENDIX : SII IN FLUCTUATING NOISE

To be able to predict speech intelligibility in the presence of background noise, the Speech Intelligibility Index or SII has been introduced in 1997 (ANSI S3.5-1997), as a revision of the Articulation Index as introduced by French and Steinberg (1947) and Kryter (1962). It calculates the total amount of speech information available to the listener, using the speech spectrum, the noise spectrum and the listener's hearing threshold as inputs. The SII has been extensively validated for stationary masking noises, but fails to predict speech intelligibility in fluctuating background. Recently, however, Rhebergen and Versfeld (2005) proposed an extension to the model, which takes the fluctuations in background noise into account. The basic principle of their approach is the partitioning of both speech and noise in small time frames. In each time frame, speech and noise are filtered into 21 critical bands to determine the 'local' spectra, which serve as inputs to determine the 'local' conventional SII. Next, the SII values of all time frames are averaged to result in the overall SII for that particular speech in noise condition. Their approach gives a good account for most existing data, producing SII's around 0.35 for normal-hearing listeners.Using their model, SII-values were calculated for the various speech-in-noise-conditions, using the speech and noise signals which were employed in our experiment. A slight modification was made concerning the filtering of the signals in critical bands, using FIR-filters with order $N_i = fs / 0.05 * f_i$, where fs is the sample frequency (44.1 kHz in our case) and f_i is the center frequency of the i-th critical band in Hertz. This assures enough resolution in the frequency domain to provide accurate filtering.

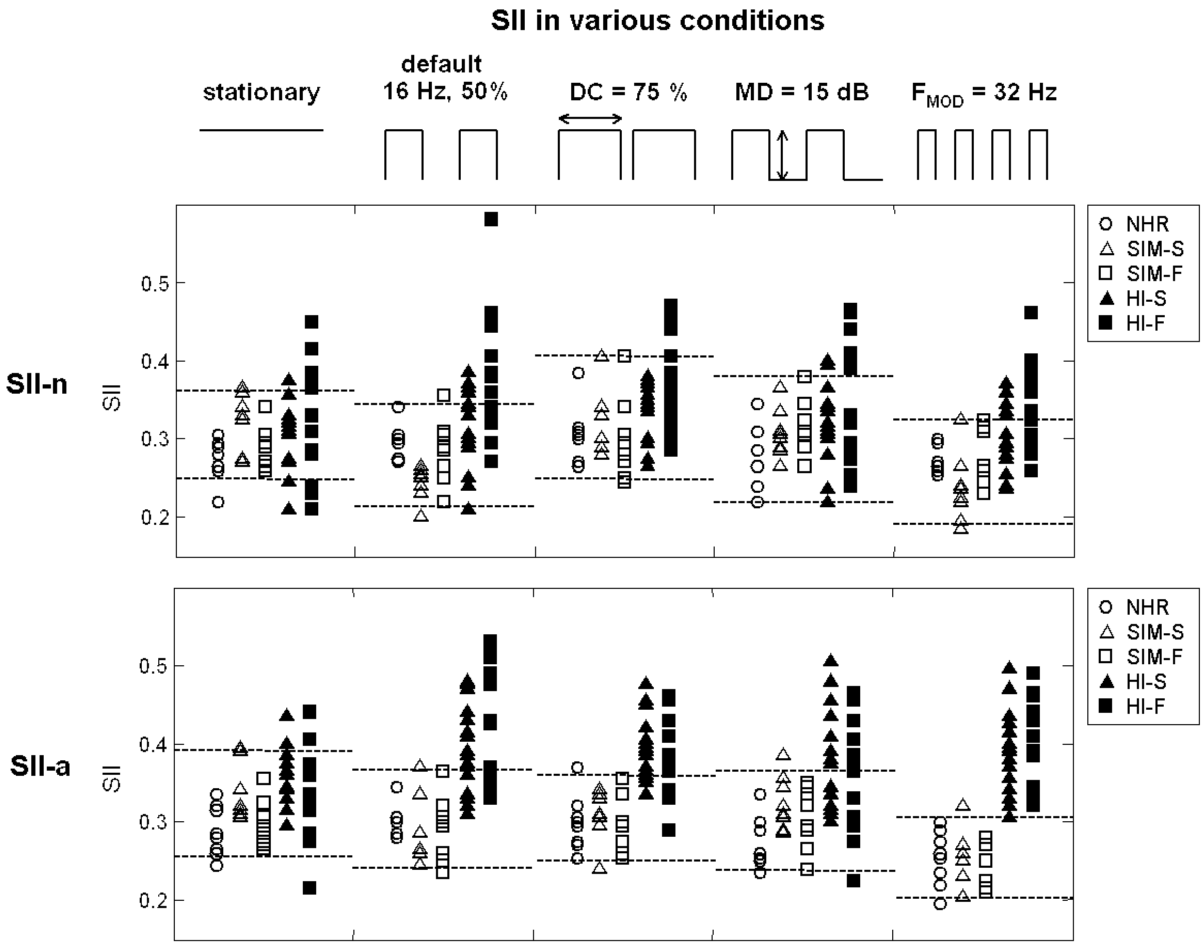

Figure 2-8 SII-values in stationary noise and in four modulated background noise conditions, in both normal (upper panel) and adapted (lower panel) spectral mode. Data are displayed for normal-hearing listeners (NHR), two groups of hearing-impaired listeners (HI-S and HI-F), and two groups of listeners with simulated hearing loss (SIM-S and SIM-F). Dashed lines represent the 5th and 95th percentiles of the open symbols. Participants in each of the five different groups have been assigned a different displacement along the x-axis for visual clarity.

This approach, however, yielded SII-values of only between 0.04 and 0.20 for normal-hearing listeners in fluctuating backgrounds, much lower than was expected. Therefore, the width of the time frames was adapted to range from 5.8 ms in the lowest band (150 Hz) to 1.9 ms in the highest band (8000 Hz), instead of the values between 35 and 9.4 ms as mentioned by Rhebergen & Versfeld (2005). These window lengths were taken from Moore et al. (1993), but scaled (with a ratio of 0.5) to be in accordance with the average temporal resolution for normal-hearing at 1000 Hz, which was measured to have a value of 4.36 ms.

Figure 2-8 shows the calculated SII-values in both adapted and normal spectral mode, in stationary noise and in four modulated background maskers. The figure shows that the calculated SII appears to be a stable measure, since listeners in both the normal-hearing and the SIM-groups display an SII of about 0.3 in most conditions. This demonstrates the power of applying the proposed SII-method in fluctuating noise, which appears to adequately incorporate problems due to only an elevated threshold.

The calculated SII for normal-hearing does, however, decrease slightly when the modulation frequency of the masker increases from the default 16 Hz to 32 Hz. Therefore, SII-values obtained in the 32-Hz-condition were not included in discussing our experimental results. Further development of the SII-model will have to aim at solving the small deviations for higher modulation frequencies.

SII_{MOD}, as used in presenting our experimental results, is thus defined as each participant's average SII over the three 16-Hz modulated masker conditions, reaching a value of around 0.3 for all normal-hearing and simulated hearing loss listeners, represented by the open symbols. As can be seen in Figure 2-8, hearing-impaired listeners generally display a larger SII, especially in adapted spectral mode. It can therefore be concluded that an SII-value of about 0.3 indicates that the listener has no auditory deficits, apart from possibly an elevated threshold, while an increased SII can serve as an indication of a supra-threshold deficit in a particular condition.

3

Auditory and non-auditory factors affecting speech reception in noise by older listeners

Erwin L.J. George, Adriana A. Zekveld, Sophia E. Kramer,
S. Theo Goverts, Joost M. Festen and Tammo Houtgast

Journal of the Acoustical Society of America (2007), **121**, 2362-2375.

Speech reception thresholds (SRTs) for sentences were determined in stationary and modulated background noise for two age-matched groups of normal-hearing (N=13) and hearing-impaired listeners (N=21). Correlations were studied between the SRT in noise and measures of auditory and non-auditory performance, after which stepwise regression analyses were performed within both groups separately. Auditory measures included the pure-tone audiogram and tests of spectral and temporal acuity. Non-auditory factors were assessed by measuring the Text Reception Threshold (TRT), a visual analogue of the SRT, in which partially masked sentences were adaptively presented. Results indicate that, for the normal-hearing group, the variance in speech reception is mainly associated with non-auditory factors, both in stationary and in modulated noise. For the hearing-impaired group, speech reception in stationary noise is mainly related to the audiogram, even when audibility effects are accounted for. In modulated noise, both auditory (temporal acuity) and non-auditory factors (TRT) contribute to explaining inter-individual differences in speech reception. Age was not a significant factor in the results. It is concluded that, under some conditions, non-auditory factors are relevant for the perception of speech in noise. Further evaluation of non-auditory factors might enable adapting the expectations from auditory rehabilitation in clinical settings.

I. INTRODUCTION

In everyday life, many older listeners have difficulties understanding speech, especially in the presence of background noise or reverberation (Plomp, 1978; Duquesnoy and Plomp, 1980; Nabelek and Robinson, 1982). A review by Plomp (1978) showed that the percentage of the population with problems in perceiving speech approximately doubles with every decade in age, from 16 percent at an age of 60 years to about 64 percent at an age of 80, to nearly everyone above 86 years of age. Although a lack of audibility, due to audiometric hearing loss or masking noise, appears to be a major component, it is generally not enough to fully account for differences in speech reception among individual listeners (Eisenberg et al., 1995; Bacon et al., 1998; Summers and Molis, 2004; Dubno et al., 2002, 2003).

Plomp (1978) formulated a model description for the speech reception threshold (SRT) based on two auditory parameters: i) hearing loss due to attenuation, related to a raised absolute hearing threshold, and ii) hearing loss due to distortion, which is considered to reflect supra-threshold deficits in hearing (Stephens, 1976; Glasberg and Moore, 1989). Examples of such supra-threshold deficits are reduced temporal and spectral auditory resolution and a loss of normal auditory compression. The inter-relationship between these deficits and their relation with the hearing threshold is still under discussion (Ludvigsen, 1985; Moore et al, 1999; Oxenham and Bacon, 2003). Both reduced temporal resolution and reduced spectral resolution are thought to adversely affect speech perception in noise, specifically when masker levels fluctuate over time (Glasberg et al., 1987; Festen and Plomp, 1990; Glasberg and Moore, 1992; Festen, 1993; Baer and Moore, 1993, 1994; Boothroyd et al., 1996; Dubno et al., 2003; George et al., 2006). Some studies, however, indicate that even listeners with significantly broadened spectral filters still have sufficient spectral resolution to resolve the spectral cues important for speech intelligibility (Ter Keurs et al., 1993a,b).

In addition, speech reception in background noise may be affected by non-auditory processes (see, e.g., Humes, 2005). Perception of speech is a process that does not only involve the peripheral auditory organ, but also depends on information processing in the central auditory pathway and on non-auditory functions, like working memory capacity and speed of information processing (Gatehouse et al., 2003; Lunner, 2003; Hällgren, 2005). More generally said, speech reception is thought to be affected by an interaction between, on the one hand, bottom-up or 'stimulus-driven' processes, and, on the other hand, top-down or 'knowledge-driven' factors

(Goldstein, 2002). The relative contribution of auditory and non-auditory processes to speech reception is, however, still under discussion.

In the early nineties, Van Rooij et al. constructed a test battery comprising auditory, cognitive and speech perception tests (Van Rooij et al., 1989). In this study, they found a significant contribution of cognitive factors to speech perception in noise. However, based on the results of two subsequent studies in older subjects (Van Rooij and Plomp, 1990; Van Rooij and Plomp, 1992), they finally concluded that age-related differences in speech perception in noise are most likely due to differences in auditory factors, notably differences in audiometric hearing thresholds. Moreover, their data indicated that auditory and cognitive factors independently contribute to speech perception, and that the importance of cognitive factors does not change significantly with increasing age.

Results by Pichora-Fuller et al. (1995) suggest that audiometric thresholds cannot fully account for the difficulty that elderly listeners experience in understanding speech in noise. They did not find any general age-related changes in cognition either; so cognitive factors do not seem responsible for the deteriorated speech recognition. Based on their results, however, they introduced a processing model, in which auditory difficulties adversely affect speech understanding both directly, by altering the amount of correctly perceived words, and indirectly, because effortful listening consumes resources that could otherwise be allocated to cognitive processes necessary for speech understanding. Consistent with this model, it was suggested by Hällgren (2005), that the relative importance of top-down or cognitive functions increases when speech information is degraded, either by hearing-impairment, or by the presence of background noise or reverberation.

Moreover, increasing evidence exists that, averaged over groups of listeners, age can affect the processing of sounds, independently of hearing loss. (Snell, 1997; Grose et al., 2001; Dubno et al., 2002; Gordon-Salant and Fitzgibbons, 2004; Gifford and Bacon, 2005). Recently, Divenyi et al. (2005) showed that the deterioration of speech reception with age is accelerated significantly relative to the decline in audiometric measures, which may indicate an accelerating decline of central processing. This finding suggests that the effect of age on speech reception might be related to non-auditory factors, as proposed earlier by for instance Gordon-Salant and Fitzgibbons (1997) and Humes (2002).

The importance of non-auditory factors is also apparent when considering the use of language proficiency (Van Wijngaarden et al., 2002, 2004) or, as more commonly investigated, sentence context (Boothroyd and Nittrouer, 1988; Nittrouer and Boothroyd, 1990; Gordon-Salant and Fitzgibbons, 1997; Dubno et al., 2000). The relevance of context was already noticed by Warren (1970), who demonstrated that listeners can perceive missing phonemes by using the redundancies in speech at the acoustic, phonetic, phonological and / or lexical level. Grant et al. (1998) referred to this function as perceptual closure, i.e. the ability to form linguistic wholes from perceived fragments. In a subsequent study, Grant and Seitz (2000) showed that this ability is modality-aspecific and may vary substantially across hearing-impaired subjects. Moreover, their results indicate that the importance of the use of context increases under degraded listening conditions, consistent with the above mentioned processing-model by Pichora-Fuller et al. (1995).

A common test to investigate the ability to make use of sentence context is the Speech Intelligibility in Noise or SPIN test (Kalikow et al., 1977; Bilger et al., 1984; for a review, see Elliott, 1995). However, SPIN-performance does not only depend on context use, but may also be related to inter-individual differences in auditory factors. In other words, measuring modality-independent factors by means of auditory stimuli may give rise to a confounding effect between auditory and non-auditory factors.

An alternative approach to confirm the importance of non-auditory factors to speech reception is to assess the relationship between auditory speech-reception performance and visual speech-reading abilities, as performed by for instance Watson et al. (1996). They found significant correlations between overall auditory and visual performance, and suggested that this association likely reflects the shared relevance of one cognitive function in visual and auditory language comprehension.

The current experiment aims at determining differences in speech reception between normal-hearing and hearing-impaired listeners and investigating the contribution of auditory and non-auditory factors in accounting for speech intelligibility in noise. In this paper, auditory factors are defined as factors that are related to processing in the peripheral and central auditory pathways, while non-auditory factors as meant here are related to modality-aspecific central, cognitive or linguistic skills.

Both normal-hearing and sensorineural hearing-impaired, older listeners were selected to participate in the present experiment. To be able to adequately compare differences between the groups, the groups were matched according to age, thus avoiding different performance due to heterogeneity of age between the groups. Speech reception was assessed by adaptive measurements of the speech reception threshold (SRT) in stationary and modulated background noise, following the procedure as described by Plomp and Mimpen (1979). To optimize audibility at all frequencies, individual hearing thresholds were used to adapt the spectrum of the noises to reach levels equal to the estimated middle of the dynamic range for each listener, as performed earlier by George et al. (2006). Measurements in modulated noise were included because differences between normal-hearing and hearing-impaired listeners are expected to be more prominent in non-stationary maskers (see, for instance, Festen and Plomp, 1990). Moreover, non-stationary maskers are more representative for listening conditions in everyday life (Kramer et al., 1996).

Correlations will be studied between SRT in noise and measures of auditory and non-auditory performance, after which stepwise regression analyses will be performed within both groups separately. Auditory performance will be assessed by measuring each listener's pure-tone audiogram and by testing both spectral and temporal acuity (F and T), which are regarded to reflect the individual auditory-filter and temporal-window width. To assess the contribution of non-auditory factors, a visually presented test was used, which was developed by Zekveld et al. (2007b) to be as similar as possible to the adaptive SRT-measurement. The outcome of this test will be referred to as the Text Reception Threshold (TRT). In this test, everyday Dutch sentences were presented visually on a screen, partially masked by an adaptively changing masking pattern. Results by Zekveld et al. (2007b) show that there is a significant correlation between the ability to read masked text (TRT) and auditory speech-reception (SRT) in a group of normal-hearing participants. The current paper will extend these results with measurements in a group of hearing-impaired listeners and age-matched normal-hearing listeners.

II. EXPERIMENT AND METHOD

A. Participants

Thirteen normal-hearing (NH) and twenty-one sensorineural hearing-impaired (HI) listeners participated in this experiment. The hearing-impaired participants (12 females, 9 males) were patients of the audiology department of the VU University

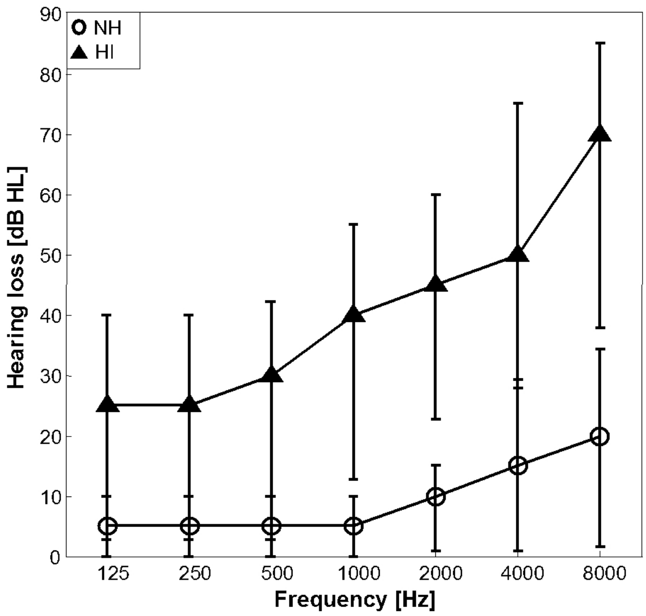

Figure 3-1 Median pure-tone hearing thresholds (*re:* ISO-389-1991) and 5th and 95th percentiles for the normal-hearing (NH, N=13) and the hearing-impaired (HI, N=21) participants.

Medical Center, selected to have a symmetrical sensorineural hearing loss, with pure-tone thresholds up to 60 dB HL and interaural threshold differences smaller than 10 dB. The age of the hearing-impaired listeners ranged from 46 to 81 years, with an average of 65.5 years. The normal-hearing listeners (8 females, 5 males) were acquaintances of the hearing-impaired participants, selected to have pure-tone hearing thresholds better than 15 dB HL at .25, .5, 1 and 2 kHz and better than 30 dB at 4 kHz. The age of the normal-hearing listeners ranged from 53 to 78 years, with an average of 63.5 years. Fig. 3-1 shows the median and spread of hearing loss for both groups of participants.

All participants were native speakers of Dutch and reported normal or corrected-to-normal vision. Their color vision was screened with Ishihara plates (Ishihara, 1989) and classified as normal.

B. Description of the tests

1. Pure-tone thresholds

Each experimental run started with the measurement of the listener's pure-tone hearing thresholds at octave frequencies between 125 and 8000 Hz, using the same apparatus as during all other measurements (see Section C). The audiogram was used to shape the spectrum of the auditory stimuli presented in the subsequent tests. In the regression analyses, only the pure-tone average (PTA) is included as a predictor variable, defined as the average pure-tone hearing threshold of the subject's best ear over octave frequencies 0.5, 1.0 and 2.0 kHz.

2. Spectral and temporal acuities (F / T)

Spectral and temporal acuities of each listener's best ear were determined by employing an adaptive measurement procedure as introduced and validated by Hilkhuysen et al. (2005). Validation was performed by measuring eighteen normal-hearing listeners. They showed auditory-filter and time-window widths which were free of noteworthy learning effects, and which varied with presentation level and frequency and corresponded to values as commonly found in the literature. The measurement procedure was also used and explained in detail in a study investigating the effects of spectral and temporal acuity on masking release for speech (George et al., 2006).

We determined spectral and temporal resolution by measuring the thresholds of short tone sweeps in spectrally or temporally modulated maskers (grids), and relating these results to the threshold in unmodulated noise. In the measurement procedure, listeners were asked to report the number of tone sweeps (zero to three) they were able to detect in i) steady-state noise without grid; ii) noise containing a spectral grid with a 50-% duty-cycle on a log-frequency scale; iii) noise containing a temporal grid with a 50-% duty-cycle. In all three noises, the tone sweeps to detect were sinusoids with a duration of 200 ms, sweeping upward over a range of 1.6 octaves centered around 1 kHz (0.57 to 1.74 kHz) at a speed of 8 octaves per second. Thus, the sweep reached its center frequency after 100 ms. The masker duration was 2.2 s, and the possible tone sweeps could start at 0.6, 1.0 or 1.4 s after masker onset. To determine detection thresholds, the level of the tone sweeps (for the steady-state noise) or the gap width of the noise-grid maskers was varied adaptively in a one-up-one-down 4-AFC-procedure (Levitt, 1971), starting above detection threshold for all listeners.

The observed masking release for the noise grids as compared to steady-state noise was used to estimate auditory-filter and time-window widths. In our experiment, both temporal and spectral acuity were determined in the frequency region around 1 kHz, at a presentation level halfway up the listener's dynamic range. Thus, signals are spectrally optimized with respect to individual hearing thresholds. This means that audibility is optimized (threshold-related effects are minimized) and the outcome measures can be assumed to be related to supra-threshold processing. A drawback of this method is, however, that the overall presentation level is different for each listener. Effects of presentation level will thus have to be considered.

Spectral resolution is known to deteriorate with increasing presentation level for normal-hearing listeners (Dubno and Schaefer, 1992; Sommers and Humes, 1993a,b), while temporal resolution is enhanced at higher levels (Jesteadt et al., 1982; Fitzgibbons, 1983; Fitzgibbons and Gordon-Salant, 1987). In our previous study (George et al., 2006), these effects of presentation level on spectral and temporal resolution were estimated by regression lines based on measurements in a group of normal-hearing listeners. It was shown that deteriorated spectral resolution for moderately hearing-impaired listeners could almost fully be accounted for by the 'normal' effects of presentation level, indicating that deteriorated spectral resolution does not qualify as an actual supra-threshold deficit. In contrast, level-corrected temporal resolution was deteriorated for most hearing-impaired listeners, indicating that it is an actual supra-threshold deficit. The deviation of measured temporal and spectral resolution from the regression lines was suggested to be the most appropriate measure for the actual amount of supra-threshold deficits of a specific listener. Therefore, in the current study, spectral and temporal acuities are corrected for presentation level in the same way.

Both level-corrected spectral and temporal acuities are included as predictor variables in the regression analyses. They are denoted by 'ΔF' and 'ΔT', where the use of the capital Greek Delta indicates that effects related to presentation level have been accounted for, as described by George et al. (2006), such that larger-than-normal values can be considered to reflect deteriorated supra-threshold processing.

3. Speech Reception Threshold (SRT)

SRT-measurements were performed using a simple adaptive one-up-one-down procedure as described by Plomp and Mimpen (1979), in stationary background noise, as well as in a masker modulated with a 16-Hz square wave with a duty-cycle of 50%. The long-term average spectra of the two maskers were the same. The appropriate masker and a list of thirteen Dutch everyday sentences were presented monaurally to the listener's best ear, sentence by sentence. Sentences were read by a female speaker (Plomp and Mimpen, 1979) and were unknown to the listener. The long-term rms-level of the masker was kept fixed, while the speech rms-level was varied adaptively to estimate the SRT, i.e. the speech-to-noise-ratio at which 50% of the sentences are reproduced without error. In each condition, the first sentence was presented at a level below threshold and repeated, at 4-dB higher levels with each repetition, until the listener was able to reproduce it correctly. The remaining twelve sentences were presented only once, following the adaptive procedure, with a 2-dB step size. The SRT was estimated as the average signal-to-noise level of sentences number 5 to 14. The fourteenth sentence was not presented, but its level was determined by the response to the thirteenth sentence.

To optimize audibility at all frequencies, individual hearing thresholds were used to adapt the spectrum of both stationary and modulated noise to reach octave masker levels equal to the estimated middle of the dynamic range for each listener. The lower limit of the dynamic range was chosen to be the individual pure-tone threshold, while the upper limit is the uncomfortable loudness level (UCL), chosen at 115 dB SPL for all listeners. Because pure-tone thresholds were only measured at octave frequencies from 125 and 8000 Hz, intermediate threshold levels were interpolated. The overall rms-level of each of the two maskers, averaged over participants, was 70.9 dB(A) for the normal-hearing group and 92.6 dB(A) for the hearing-impaired group, as can be derived from the audiograms given in Fig. 3-1. The shape of the speech spectrum was modified to be the same as that of the noise, while the level of the speech was varied, following the adaptive procedure as described above. The bandwidth of the noise and the speech signals was restricted to frequencies between 223 and 4490 Hz.

SRT in stationary noise and SRT in modulated noise are identified as criterion variables in the regression analyses, and are denoted by 'SRT_{STAT}' and 'SRT_{MOD}', respectively.

4. Text Reception Threshold (TRT)

In order to determine whether non-auditory processes could play a role in the reception of speech, a visually presented test was used that is similar to the auditory SRT-test. This test was developed earlier by Zekveld et al. (2007b), who used it to investigate the relation between the ability to comprehend speech in noise and the ability to read masked written text in normal-hearing subjects. A list of thirteen everyday Dutch sentences, adopted from Versfeld et al. (2000), was presented visually on a computer screen, sentence by sentence. Each presented sentence was partially masked by vertical bars with a specific degree of masking, adaptively changing over sentences. The field back colour was white, text colour was red, and the colour of the mask was black. The text of each sentence appeared in a word-by-word fashion. The timing of the appearance of each word was based on the original recordings that were available for each sentence. The sentence disappeared 3.5 seconds after presentation of the last word of the sentence, while the mask remained visible until the next trial started with the presentation of a newly created mask.

The number of vertical bars in the masking pattern was determined such that for most masking percentages, each letter was partly masked and partly unmasked. Examples of sentences masked by this pattern are shown in Fig. 3-2. The exact position of the mask relative to the sentence was randomly determined.

The amount of masking was varied adaptively, following an one-up-one-down procedure similar to the SRT-test, to determine the Text Reception Threshold or TRT, defined as the amount of unmasked text needed by the subject to comprehend 50% of complete sentences correctly. Zekveld et al. (2007b) showed that changing the amount of unmasked text with a step size of 6% would yield a change in the proportion of correct responses comparable to the 2-dB steps in the SRT-test. In each condition, the first sentence was presented at a masked text percentage below threshold and was repeated, with an increased percentage of unmasked text, until the listener was able to reproduce it correctly. Also similar to the SRT, a double step size

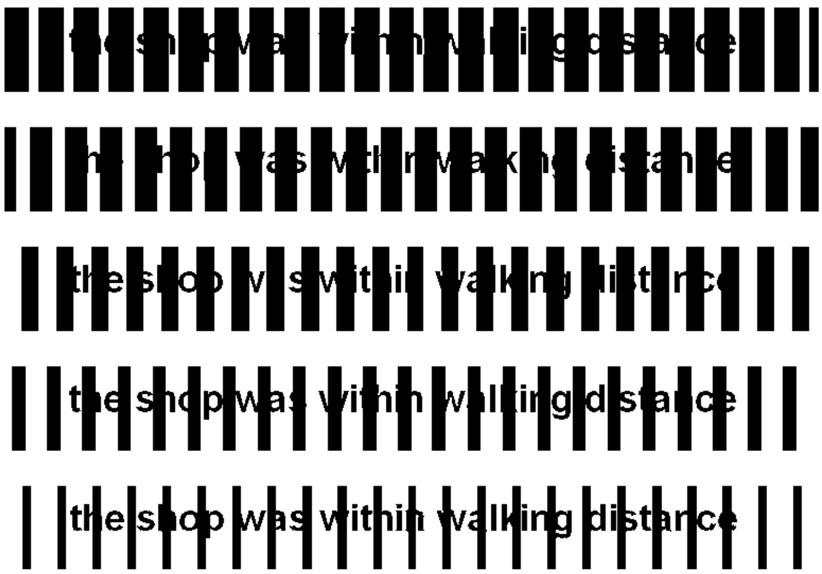

Figure 3-2 Typical stimuli to measure the Text Reception Threshold (TRT): a sentence masked by a vertical bar pattern. Between sentences, the degree of masking was adaptively varied. The field back colour was white, text colour was red, and the colour of the mask was black. The shown percentages of unmasked text are 28, 40, 52, 64 and 76 percent, respectively. For the purpose of this figure, a Dutch sentence from the lists by Versfeld et al. (2000) was translated into English. The figure was adopted from Zekveld et al. (2007).

(12%) was used for the first sentence. The remaining twelve sentences were presented only once, following an adaptive procedure with a 6% step-size. The TRT was estimated as the average percentage of unmasked text of sentences 5 to 14. It is regarded here as a general measure of modality-aspecific cognitive and linguistic skills contributing to the perception of partially masked sentences. The TRT is included as a predictor variable in the regression analyses.

C. Instrumentation and general procedure

The experiment was run on a Dell personal computer, equiped with a Creative Labs Audigy external sound device and Beyer Dynamic DT48 headphones. Sound calibrations were performed with a Brüel & Kjær Artificial Ear (type 4152) and a Brüel & Kjær 2260 Observer conform ISO 389 (1991). All measurements were performed while the listener and the investigator were seated in a sound-insulated room. Spectral shaping of auditory signals was performed by using individual thresholds as inputs via a 1024-point windowed FIR filter. This filter also corrected the headphones frequency response and restricted the bandwidth of auditory signals to frequencies between 223 and 4490 Hz.

All measurements were performed following a test-retest design in a single session, interrupted by several small breaks. The test and retest blocks each included the measurement of three TRT, two SRT_{MOD} and single SRT_{STAT}, F and T. The TRT and the SRT_{MOD} measurements were performed more than once, both in test and retest, to improve reliability. Test and retest outcomes were averaged. The order in which the tests were presented was fixed. A session always started with the measurement of the audiogram and color vision screening. Auditory measurements were conducted monaurally, using the participant's best ear, which was chosen according to his or her audiogram (PTA), or, in case of equal PTA's, personal preference in telephone conversation.

D. Statistical analysis

Statistical analyses were performed using SPSS for Windows, release 11.0.1. Kolmogorov-Smirnov tests were performed to check the normality of variables, which showed that all variables were normally-distributed within each group. Considering the expected non-equal variances in the two groups, two-tailed t-tests assuming non-equal variances were used to examine group differences between the test-outcomes of the normal-hearing and the hearing-impaired participants. Correlation coefficients were calculated for each of the groups separately, to investigate which predictor variables are significantly associated with speech reception in stationary and modulated noise. Although it is common practice to scale down the significances when performing multiple comparisons or correlations (Miller, 1981), it was decided to adopt a tolerant criterion for significance, considering the exploratory nature of our study. All effects reaching *p*-values below 0.05 are indicated by asterisks.

Finally, to account for predictor cross-correlations, stepwise multiple regression analyses were performed for both groups to investigate which predictor variables could most effectively account for inter-subject variance in speech reception. The variables PTA, ΔF, ΔT, TRT and age were included as predictor variables in these analyses, while SRT_{STAT} and SRT_{MOD} were included as criterion variables. The percentages of explained variance (R^2) reported below have already been corrected for the available degrees of freedom.

III. RESULTS

A. Differences between groups

Table 3-I displays group averages, standard deviations, reliabilities, standard errors of measurement and t-statistics of the test-outcomes for normal-hearing and hearing-impaired participants. The test-outcomes are ordered by origin (auditory or non-auditory). To investigate to which extent the multiple measurements of both the TRT and the SRT_{MOD} tests increase reliability and significance of group differences, Table 3-I also shows data when these variables would have been measured only once in test and retest, like all other test outcomes.

Results of the t-test show that group differences for the SRT in modulated noise (SRT_{MOD}) are larger than for the SRT in stationary noise (SRT_{STAT}). Even when SRT_{MOD} is based on only one test-retest measurement (SRT_{MOD}**), the difference in speech reception between the groups appears to be more prominent in modulated noise compared to stationary noise, as indicated by the larger t-value. To statistically test this hypothesis, a two (group) by two (background noise) repeated-measures ANOVA was performed on SRT_{STAT} and SRT_{MOD}**, which showed significant effects of group (p = .004) and background noise (p < .001), but also a significant interaction between group and background noise (p = .01). This interaction indicates that the difference in speech reception between the groups is significantly larger in modulated noise compared to stationary noise. Thus, measuring the SRT in modulated noise instead of in stationary noise indeed increases the ability to discriminate between normal-hearing and hearing-impaired participants.

In addition, it can be concluded from the t-statistics that the performances of both groups differ significantly for auditory tests, like temporal acuity (ΔT) and audiogram (PTA). The latter is not surprising, since the groups were selected on the basis of their differences in audiometric thresholds. In contrast, spectral acuity (ΔF) does not appear to differentiate between normal-hearing and hearing-impaired listeners, as shown before by George et al. (2006). Moreover, Table 3-I shows that the normal-hearing and hearing-impaired groups do not perform significantly different on the TRT-test.

The fact that differences between the two groups are mainly auditory in nature indicates that differences in speech reception between the groups are also likely to be

			NH (N=13)				HI (N=21)					
		unit	M	S	r_{tt}	SEM	M	S	r_{tt}	SEM	t	p
Auditory	ΔF	Hz	212.9	22.8	.60	14.4	220.5	79.3	.84	32.0	0.408	.687
	ΔT	ms	3.59	0.85	.98	0.12	6.98	3.07	.92	0.84	4.776	< .001
	PTA	dB HL	6.7	3.0	-	-	34.4	12.0	-	-	10.041	< .001
	SRT_{STAT}	dB SNR	-3.2	1.1	.60	0.7	-1.4	2.3	.83	0.9	2.966	.006
	SRT_{MOD}	dB SNR	-14.0	2.6	.88	0.9	-9.0	4.4	.96	0.9	4.139	< .001
	SRT_{MOD}**	dB SNR	-13.1	3.3	.82	1.4	-8.2	5.2	.88	1.8	3.408	.002
Non-auditory	age	years	63.5	9.3	-	-	65.5	9.9	-	-	0.576	.569
	TRT	% text	58.2	4.3	.88	1.5	58.8	3.3	.87	1.2	0.427	.674
	TRT**	% text	58.9	4.7	.83	1.9	59.7	4.2	.79	1.9	0.510	.615

*** based on only one measurement in test and retest, instead of two (SRT_{MOD}) or three (TRT)*

Table 3-I Group averages (M), standard deviations (S), reliabilities (r_{tt}), standard errors of measurement (SEM) and two-tailed t-statistics of the test-outcomes for the normal-hearing (NH) and the hearing-impaired (HI) participants. Test reliabilities r_{tt} have been calculated from test-retest correlations r_{tr} using the Spearman-Brown formula: $r_{tt} = 2*r_{tr} / (1+r_{tr})$, cf. Nunnally, 1967. SEM is defined as $S * \sqrt{(1-r_{tt})}$. Group differences are considered significant if $p < 0.05$.

mainly governed by auditory factors. To investigate whether non-auditory factors nevertheless are associated with inter-individual variance in speech reception in noise, results were investigated within both groups separately, thus partializing out the group-effect.

B. Correlation analyses within groups

Table 3-II gives product-moment correlations between the tested predictor variables and SRT in stationary noise (SRT_{STAT}) and in modulated noise (SRT_{MOD}) for both the normal-hearing and the hearing-impaired participants. Age is included as an extra predictor variable.

It can be seen in Table 3-II that, for the normal-hearing listeners, non-auditory factors have the largest association with speech intelligibility, especially in modulated noise. In contrast, for the hearing-impaired group, Table 3-II shows that non-auditory factors do not appear to be significantly associated with individual differences in speech reception. Instead, mainly auditory factors, specifically temporal acuity (ΔT) and audiogram (PTA), appear to be related to speech intelligibility in this group, both in stationary and non-stationary noise. It should be noted, however, that the differences in ranges of auditory and non-auditory measures between both groups contribute to the observed differences in correlations, as will be discussed in Section A of the Discussion.

		NH (N=13)		**HI (N=21)**	
		SRT_{STAT}	**SRT_{MOD}**	**SRT_{STAT}**	**SRT_{MOD}**
Auditory	ΔF	.12	.22	- .05	.13
	ΔT	- .26	- .09	**.52** $^{(*)}$	**.73** $^{(***)}$
	PTA	.39	.48	**.71** $^{(***)}$	**.73** $^{(***)}$
Non-auditory	age	.37	.43	.29	.39
	TRT	**.61** $^{(*)}$	**.80** $^{(***)}$	.34	.42

Table 3-II Product-moment correlations between the predictor variables and SRT in stationary noise (SRT_{STAT}) and SRT in modulated noise (SRT_{MOD}). All correlations are calculated separately for the normal-hearing (NH) and the hearing-impaired (HI) participants. Significant correlations are displayed in bold, p-values are denoted by asterisks: $^{(*)}$ $p < 0.05$; $^{(**)}$ $p < 0.01$; $^{(***)}$ $p < 0.001$.

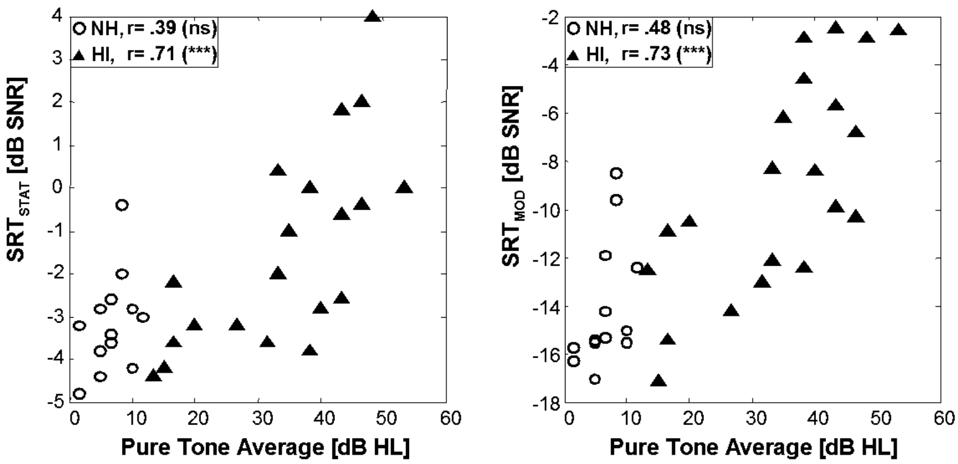

Figure 3-3 SRT in stationary noise (SRT$_{STAT}$, left) and in 16-Hz block-modulated noise (SRT$_{MOD}$, right) as a function of pure-tone average (PTA), for the normal-hearing (NH, N=13) and hearing-impaired (HI, N=21) participants. Also shown are the correlation coefficients and their significances: (ns) non-significant; (*) $p < 0.05$; (**) $p < 0.01$;

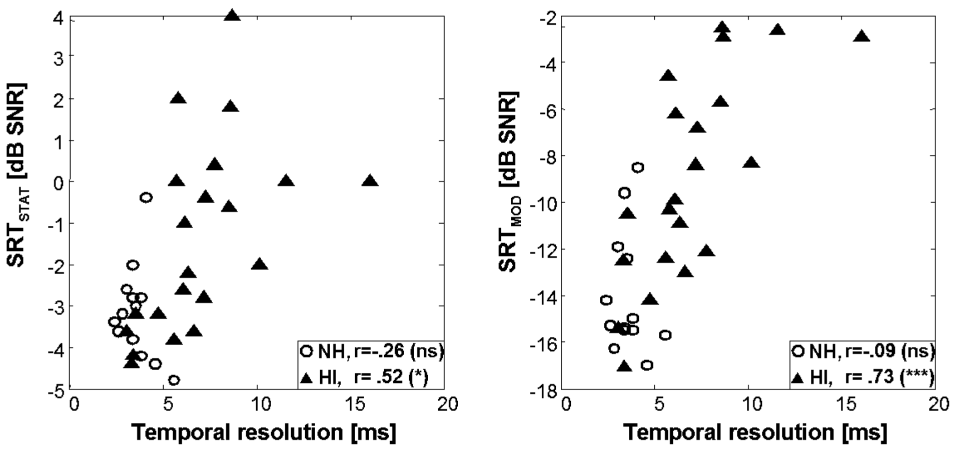

Figure 3-4 SRT in stationary noise (SRT$_{STAT}$, left) and in 16-Hz block-modulated noise (SRT$_{MOD}$, right) as a function of temporal resolution (ΔT), for the normal-hearing (NH, N=13) and the hearing-impaired (HI, N=21) participants. Correlations and significances indicated as in Fig. 3-3.

Figure 3-3 and 3-4 show, for the SRT in stationary and in modulated noise, the relationship with the audiogram (PTA) and temporal acuity (ΔT), respectively. These figures illustrate that audiogram and temporal resolution can differentiate between the normal-hearing and hearing-impaired groups, as indicated by the t-statistics in Table 3-I. However, they also confirm that auditory factors cannot account for all of the variance in SRT, specifically within the normal-hearing group. The correlation analysis shows that part of this variance may be related to inter-individual differences in non-auditory factors, as measured by the TRT.

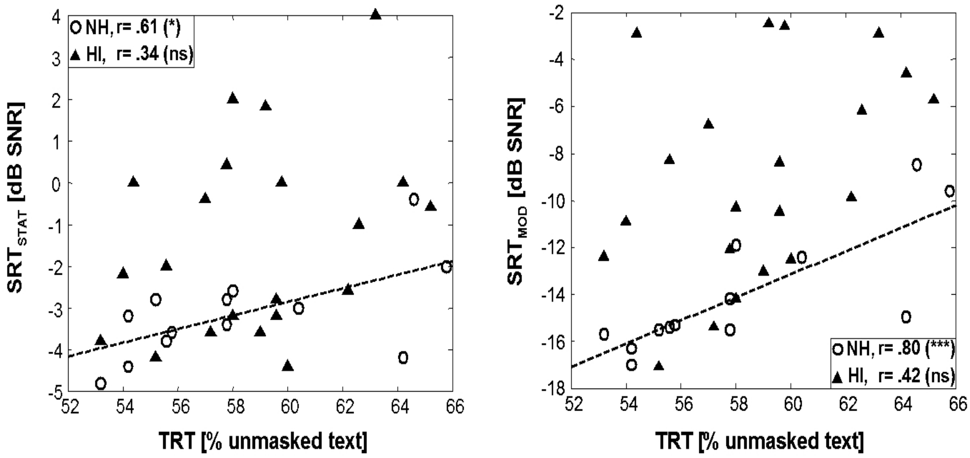

Figure 3-5 SRT in stationary noise (SRT_{STAT}, left) and in 16-Hz block-modulated noise (SRT_{MOD}, right) as a function of Text Reception Threshold for text masked with a vertical bar pattern (TRT), for the normal-hearing (NH, N=13) and the hearing-impaired (HI, N=21) participants. The dashed lines are linear regression lines fitted on the data of the normal-hearing participants. Correlations and significances indicated as in Fig. 3-3.

To further investigate the relation between the TRT and SRT in stationary and modulated noise, SRT has been plotted as a function of the TRT for both groups in Fig. 3-5. It can be seen that the range (i.e. average and variance) of the TRT, as given in Table 3-I, is the same for the normal-hearing and hearing-impaired listeners. This illustrates the results from the t-test performed earlier: the groups can not be distinguished on the basis of the TRT. Fig. 3-5 will be rediscussed later in the light of the results of the stepwise multiple regression analysis.

Finally, it can be seen in Figures 3-3 to 3-5 that the ranges of both SRT_{STAT} and SRT_{MOD} are different for the hearing-impaired group and for the normal-hearing group. This illustrates that the hearing-impaired group performs significantly worse than the normal-hearing group on both SRTs, as shown by the t-statistics in Table 3.I. The difference in SRT between both groups appears more prominent in modulated noise, though.

C. Stepwise multiple regression within groups

As mentioned earlier, the above presented correlation analysis on speech reception is not ideal. It gives a distorted picture, because the predictors are cross-correlated, as can be seen in Table 3-III. This possibly leads to induced correlations between a predictor and speech intelligibility. To control for these cross-correlations, stepwise multiple regression analyses were performed, again for each group of participants separately.

		Auditory			Non-auditory	
		ΔF	ΔT	PTA	age	TRT
Auditory	ΔF	-				
	ΔT	.42	-			
	PTA	- .28	**.59(**)**	-		
Non-auditory	age	- .22	.09	.42	-	
	TRT	- .35	- .08	.35	.28	-

Table 3-III Product-moment cross-correlations between the predictor variables. Only data from the hearing-impaired participants (N=21) were included in the calculations. Significant correlations are displayed in bold, p-values are denoted by asterisks, indicated as in Table 3-II.

For the normal-hearing group, results from the stepwise multiple regression analysis show that the TRT is the predictor variable that contributes most to explaining variance in speech reception, both in stationary and in modulated noise. On its own, the TRT accounts for 31% of the inter-subject variance in speech reception in stationary noise and for 60% of the variance in speech reception in modulated noise, as already indicated by the correlations between TRT and SRT displayed in Table 3-II. When the TRT is included in the model, no other predictor significantly explains variance over and above the variance explained by the TRT ($p > .46$ in both cases).

For the hearing-impaired group, results from the stepwise multiple regression analysis are displayed in Table 3-IV. In stationary noise, only PTA significantly contributes to explaining differences in speech reception, accounting for 47% of the inter-subject variance. When PTA is included in the model, no other predictors are significantly correlated with the residual ($p > .34$). In modulated noise, mainly temporal acuity (ΔT) contributes to explaining variance in speech intelligibility for hearing-impaired participants, accounting for 48% of the variance on its own. The TRT explains only 9% of the variance ($p = .06$) when it is the only predictor included, but becomes a significant term ($p = .0006$) when temporal acuity is included in the model first. Together, ΔT and the TRT explain 73% of the variance in speech intelligibility in modulated noise. When they are both included in the model, no other predictors significantly contribute anymore ($p > .09$).

	SRT$_{STAT}$			SRT$_{MOD}$		
	predictor	R^2	cum. R^2	predictor	R^2	cum. R^2
step 1	PTA	.47	.47	ΔT	.48	.48
step 2	-	-	-	TRT	.09	.73

Table 3-IV Results from the stepwise multiple regression analyses: significant contributors to explaining the variance in SRT in stationary noise (SRT$_{STAT}$) and SRT in modulated noise (SRT$_{MOD}$). The analysis was performed only on the data from the hearing-impaired participants (N=21). Shown are the successive contributing predictor variables and the percentage of the variance that they significantly account for when included in the model, either on their own (R^2) or cumulatively combined (cum. R^2). Both measures are corrected for the available degrees of freedom. All shown models have a significance $p < 0.001$.

Thus, the regression analysis shows that a non-auditory component, as measured by the TRT, does appear to significantly contribute to explaining variance in SRT$_{MOD}$ for the hearing-impaired subjects, in contrast with what was to be expected from the correlations shown in Table 3-II. Even though the correlation between the TRT and SRT$_{MOD}$ (r = .42) is non-significant, the TRT is significantly associated with SRT$_{MOD}$ when the effect of temporal acuity is partialized out by the first step in the regression analysis.

In summary, there is a difference in the relative contribution of factors accounting for variance in speech reception between normal-hearing and hearing-impaired listeners. When there are no or only minor auditory problems, as in the normal-hearing participants, variance in speech reception in noise is mainly governed by non-auditory factors. However, when variance in auditory processing increases as a consequence of deterioration in the auditory system, inter-individual differences in speech reception appear to be governed by both auditory and non-auditory factors.

IV. DISCUSSION

A. Differences between groups

Our experimental results indicate that non-auditory factors, as measured by the TRT, are the most important source of variance in speech reception for the normal-hearing listeners. For the hearing-impaired listeners, both auditory and non-auditory factors appear to influence inter-subject differences in speech reception.

This observed difference in contribution of auditory and non-auditory factors can be explained by the differences in ranges of the auditory and non-auditory factors between the two groups. Participants in the normal-hearing group all perform reasonably well on auditory factors, i.e. the range of for instance PTA and ΔT is only fairly small compared to the range in the hearing-impaired group. This difference in ranges contributes to the observed relatively small correlation between auditory factors and SRT for normal-hearing participants. Thus, the relatively small range of auditory factors for the normal-hearing group, compared to the larger range of auditory factors in the hearing-impaired group, may be considered the main reason for the apparent difference in the relative contribution of auditory and non-auditory factors between both groups.

In contrast, the range of the non-auditory TRT is about the same for both groups, making the TRT more likely to contribute to variance in both groups. Indeed, the TRT contributes to explaining variance in SRT in both groups, especially in modulated noise. Nevertheless, the observed correlation between TRT and SRT_{MOD} is smaller for the hearing-impaired group, due to the influence of auditory factors, which explain the main part of the variance. Thus, the relative contribution of the TRT to speech reception is larger in the normal-hearing group (explaining 60% of the variance) than in the hearing-impaired group (25% after temporal acuity is partialized out in the first step of the regression analysis). However, these relative contributions can be expressed in absolute terms by multiplying them with the total amount of unexplained variance (i.e. the squared standard deviation) within a group. In absolute terms, then, the TRT explains a similar amount of variance in SRT_{MOD} in both groups (4.2 dB^2 in the normal-hearing and 4.9 dB^2 in the hearing-impaired group). This finding indicates that non-auditory factors affect speech reception in modulated noise, independent of the amount of hearing loss.

B. Differences between stationary and modulated noise

Considering the results within the hearing-impaired group, there seems to be a fundamental difference between the factors that explain variance in speech intelligibility in either stationary or modulated noise. The present results show that speech reception in stationary noise is mainly governed by auditory factors, while both auditory and non-auditory factors account for inter-subject variance in modulated noise.

In stationary noise, PTA is the main factor accounting for inter-subject variance in speech intelligibility. This is in agreement with results by Van Rooij and Plomp (1992), who concluded that the audiogram is the most adequate predictor of speech reception in stationary noise. In modulated noise, temporal acuity (ΔT) and TRT appear to be the main factors explaining inter-subject differences in speech reception. This does not mean that PTA does not correlate with SRT in modulated noise, as can be seen in Table 3-II (r = .73). It does mean, however, that PTA does not explain as much variance in SRT_{MOD} as temporal acuity and the TRT do together, and that, in the regression analysis, the effect of PTA is covered by these variables.

These findings suggest that PTA is an estimate for general auditory performance and, as such, is related to speech reception in both stationary and modulated noise. Measuring speech reception in modulated noise, showing larger inter-individual differences, apparently enables specification of the variance in speech reception into temporal auditory processing (temporal acuity) and non-auditory factors (TRT). Moreover, measuring the SRT in modulated noise appears to increase discrimination between normal-hearing and hearing-impaired participants, as indicated by the t-statistics in Table 3-I. Therefore, modulated noise may be preferred over stationary noise to measure speech reception for clinical purposes, specifically because non-stationary backgrounds are more common in everyday situations (Kramer et al., 1996).

C. Relation between speech reception (SRT), PTA and presentation level

In the speech reception tasks in the current experiment, individual hearing thresholds were used to adapt the spectrum of the background noise and the speech signal, i.e. the noise spectrum was fixed halfway up the dynamical range for each participant. This method optimizes audibility at all frequencies for each participant, but, consequently, gives rise to inter-individual spectrum and level-differences. As mentioned above, the average rms-level of each of the two maskers was 70.9 dB(A) for the normal-hearing group and 92.6 dB(A) for the hearing-impaired group. Using the SRT to quantify the subjects' ability to perceive speech in noise does not take these differences between listener groups into account.

Moreover, optimizing audibility for each listener has as a direct consequence that presentation level and hearing threshold are related, and that their effects on speech reception cannot be fully distinguished. This means that part of the correlation between the audiogram (or PTA) and the SRT, as reported above, might be attributed to the effects of presentation level. Indeed, overall presentation level, expressed in dB(A), is significantly related to SRT in the hearing-impaired group, both in stationary (r = .44, $p = .05$) and in modulated noise (r = .57, $p = .007$).

A measure of speech intelligibility performance which is able to handle inter-subject audiogram and spectrum differences is the Speech Intelligibility Index or SII (ANSI S3.5-1997), which gives an estimate of the amount of speech information available in a certain condition, using the individual's audiogram and the signal and masker spectrum levels as inputs. In addition to accounting for threshold and spectrum-differences, the SII-model also includes a level distortion factor, which takes the deterioration of speech recognition at higher presentation levels into account. An SII of about 0.30 to 0.35 is commonly considered to be enough to reach 50% speech intelligibility for normal-hearing listeners. Hearing-impaired listeners generally need more speech information, which is regarded as the result of less efficient processing of the information due to supra-threshold deficits.

The variables SRT_{STAT} and SRT_{MOD} were tranferred to SII-values by using the model as introduced and validated by Rhebergen & Versfeld (2005), which was also applied in our earlier study (George et al. 2006). Multiple stepwise regression analyses were performed as before, with SII_{STAT} and SII_{MOD} as dependent variables, to investigate which predictor variables could most effectively account for inter-subject variance in SII. For easy comparison, the amounts of explained variance in SRT, as obtained before in Section C of the Results, will be repeated below between brackets.

Results of the stepwise regression analyses on SII-values show that, in the normal-hearing group, the TRT is the predictor variable that contributes most to accounting for variance in SII-values. On its own, the TRT accounts for 24% [31%] of the inter-subject variance in SII_{STAT} ($r = .55$, $p = 0.05$) and for 58% [60%] of the variance in SII_{MOD} ($r = .78$, $p = 0.002$). When the TRT is included in the model, no other predictor significantly explains variance over and above the variance explained by the TRT ($p > .43$ in both cases).

For the hearing-impaired group, only PTA significantly contributes to inter-subject differences in SII, accounting for 21% [47%] of the variance in SII_{STAT} ($r = .50$, $p = 0.02$). When PTA is included in the model, no other predictors are significantly correlated with the residual ($p > .26$). In modulated noise, temporal acuity (ΔT) contributes most to explaining variance in speech reception in the hearing-impaired group, accounting for 49% [48%] of the inter-subject variance in SII_{MOD} ($r = .72$, $p < 0.001$). The TRT is not significantly related to SII_{MOD} ($p = .16$), but becomes a significant term ($p = .01$) when temporal acuity is included in the model first. Together, ΔT and the TRT explain 62% [73%] of the variance in SII in modulated noise. When they are both included in the model, no other predictors significantly contribute anymore ($p > .30$).

The results of the regression analyses on SII-values are comparable to the earlier obtained results on SRT-values. The largest difference is the reduced amount of variance explained by PTA in the hearing-impaired group in stationary noise. This was to be expected, since it is a direct consequence of taking inter-individual audibility, spectrum and level differences into account. Even though the SII may underestimate the deteriorating effects of level on speech reception, as argued by Studebaker et al. (1999), the correlation between overall presentation level and speech reception, as observed before in the hearing-impaired group, is not significant anymore when the SII-values are considered (r = .38, $p = .09$ in stationary noise; r = .32, $p = .15$ in modulated noise).

Within the current set of predictor variables, PTA is still the best (least poor) predictor of speech reception (SII) in stationary noise for hearing-impaired listeners, even though level and audibility differencese have been taken into account. This confirms our earlier suggestion that the correlation between PTA and speech reception is not related to audibility differences. Instead, it can be understood by considering PTA as a good estimate for general auditory performance: the development of hearing loss (PTA) generally accompanies deterioration of supra-threshold processing, and vice-versa. As shown in Table 3-III, PTA is indeed significantly related to temporal acuity, and it is not unlikely that a larger PTA also reflects deteriorated intensity coding, recruitment (Stephens, 1976) or the loss of normal auditory compression. In this light, it is understandable that PTA relates to SRT in stationary noise, even when audibility effects are accounted for.

In summary, these SII-results corroborate to our earlier conclusions that, in the normal-hearing group, variance in speech reception in noise is mainly governed by non-auditory factors. In the hearing-impaired group, inter-individual differences in speech reception appear to be governed by both auditory and non-auditory factors, especially in modulated noise.

D. Relation between speech reception (SRT) and ΔF, ΔT and age

The observed non-contribution of deteriorated spectral resolution (ΔF) to variance in speech reception does not appear to be consistent with literature (Patterson et al., 1982; Noordhoek et al., 2001). This may be explained, however, by the fact that deteriorated spectral acuity for mild to moderate hearing-impaired listeners may be largely accounted for by increased presentation level, as demonstrated by George et al. (2006). When presentation level is accounted for, differences in spectral resolution between normal-hearing and hearing-impaired participants are only minor, making spectral acuity a poor candidate to explain variance in speech reception (see the non-significant correlations in Table 3-II). In contrast, Table 3-II shows that reduced temporal resolution (ΔT) is significantly related to speech reception in both stationary and modulated noise. In stationary noise, there is a significant correlation between ΔT and SRT, but the regression analysis shows that this effect disappears after taking PTA into account, while it is just the other way around in modulated noise. This can be explained by the fact that, in modulated noise, the effect of deteriorated temporal resolution is more prominent. That is, sufficient temporal acuity is necessary for a listener to take advantage of the relatively silent periods or gaps in modulated noise (Glasberg et al., 1987; Festen and Plomp, 1990; Peters et al., 1998; Snell et al., 2002).

Age does not appear to be significantly related to speech reception in noise, as implied by the correlations in Table 3-II and by results of the regression analyses. This finding does not appear to be in line with earlier findings, which did report a detrimental effect of age on speech reception (e.g. Gustafsson and Arlinger, 1993; Snell et al., 2002). One possible explanation for this non-contribution of age might be that age-related deficits are most apparent in complex auditory tasks or backgrounds (Gordon-Salant, 1987; Souza and Turner, 1994; Pichora-Fuller and Souza, 2003). That is, perhaps our listening conditions weren't challenging enough for the participants. However, this explanation seems unlikely, since the adaptive procedure used in the speech reception task varies the signal-to-noise ratio around the point at which the listener reaches 50% sentence intelligibility, thus avoiding 'easy listening conditions'. A more likely explanation is found by suggestions in literature (Pichora-Fuller et al., 1995; Watson et al., 1996; Gordon-Salant and Fitzgibbons, 1997; Gatehouse et al., 2003) that the effect of age on speech reception in noise may be mediated by the influence of cognitive effects on hearing ability. In fact, in the current experiment, age was significantly related to SRT_{MOD} (p = 0.012) when temporal acuity was partialized out by the first step in the regression analysis for hearing-impaired listeners. The correlation between TRT and SRT_{MOD}, however, was larger (p = 0.0006), giving rise to a larger cumulatively explained amount of variance. When TRT is included in the regression model, the contribution of age is not significant anymore. This implies that the inclusion of TRT in the model reduces age-related variance in SRT_{MOD}. Thus, even though the relationship between age and TRT was not statistically significant, they share a common component related to speech reception, which is better expressed in terms of non-auditory factors (TRT) than in terms of age.

Finally, it should be noted that the observed percentages of variance accounted for (R^2), as mentioned in the Results Section, may be underestimates, considering the non-unity reliabilities of the predictor variables, as shown in Table 3-I. The R^2 were substantial, although only a relatively small number of participants was included in the current experiment, which may indicate that there may be little systematic variance left to be explained over and above the variance explained by the predictor variables included in the present study.

E. Relation between speech reception (SRT) and text reception (TRT)

The TRT does not only contribute significantly to explaining variance in SRT_{MOD} in the hearing-impaired group, but is also the main source of variance in speech reception in the normal-hearing group. The obtained correlation in this group between SRT_{MOD} and TRT ($r = .80$) suggests that the perception of speech in non-stationary noise and of masked text depend partly on the same modality-aspecific processing mechanisms. This finding confirms the results of Zekveld et al. (2007b), who observed this relationship between SRT and TRT in a normal-hearing group with a wider age range. Moreover, this result is in line with results of, for instance, Amitay et al. (2002), who demonstrated the relationship between spoken and written language comprehension in a study on reading disabilities. Consistent with the suggestions of Grant et al. (1998) and Zekveld et al. (2007b), we suggest that this shared component reflects a modality-independent skill to perform perceptual closure, i.e. the process by which the masked portions of a stimulus are completed in order to identify the object (Snodgrass & Kinjo, 1998). Specifically, the closure as meant here is related to the extent to which listeners restore missing phonemes or words by using sentence context, i.e. the redundancies in speech at the acoustic, phonetic, phonological and / or lexical level (Warren, 1970).

Consistent with findings by Grant and Seitz (2000), it was observed that this ability, as expressed by the TRT, may vary substantially across hearing-impaired subjects. Apparently, some hearing-impaired subjects are better at using sentence context for speech recognition than others. This may enable the application of the TRT for clinical purposes.

To clarify this point, Fig. 3-5 should be reconsidered. It illustrates the results from the stepwise regression analyses, showing that for both the normal-hearing and the hearing-impaired participants, the deteriorating SRTs in the two noise conditions appear to be associated, to some extent, with deteriorating TRT. For an individual normal-hearing participant, the SRT is mainly governed by non-auditory factors, that also govern the TRT. Hence, for normal-hearing participants, the relationship between SRT and TRT is relatively strong. The dotted linear regression lines represent this relationship in Fig. 3-5. For hearing-impaired participants, the relation between the TRT and SRT in noise appears less straight-forward. The normal-hearing regression lines might be considered as a 'limit of performance' or 'baseline' for hearing-impaired subjects. When a data-point is close to the regression line, the SRT can be considered essentially normal, given the TRT, or, to put it differently, the SRT-score can be considered to be solely related to non-auditory factors. An individual's vertical deviation from the normal-hearing regression line may then be thought of as reflecting the relative contribution of deteriorated auditory factors to the elevated SRT. The stepwise regression results indicate that, in modulated noise, this deviation is related to deteriorated temporal acuity (ΔT).

Thus, the combined measurements of SRT_{MOD} and TRT makes it possible to estimate the relative contribution of auditory and non-auditory factors to speech recognition in noise. This may enable the clinical examiner to determine in part the origin (auditory or non-auditory) of deteriorated speech reception, using the TRT-test as an additional diagnostic clinical tool. A listener with a relatively poor TRT-score will be less able to use context information to improve sentence readability or intelligibility. Thus, to reach equal speech intelligibility, this listener will need a relatively higher signal-to-noise ratio compared to a listener with a better TRT-score. The expectations concerning the benefit from auditory rehabilitation by, for instance, hearing-aids, can then be adapted likewise (see Pichora-Fuller and Souza, 2003).

The success of clinical application of the combination of TRT and SRT_{MOD} does, however, depend on the assumption that the relative contributions of auditory and non-auditory factors to speech reception do not change with deteriorating auditory performance or age. Differently said, it is assumed that auditory and non-auditory factors are independently used to process speech, such that the deterioration of auditory processing does not change the relative contribution of non-auditory factors to speech reception.

Although it has been suggested by Pichora-Fuller et al. (1995) that auditory and non-auditory (cognitive) functions are linked in a processing model, there is as yet no reason to doubt this assumption. The fact that the TRT accounts for the major part of the variance in speech-reception in the normal-hearing group suggests that the TRT is capable of measuring modality-aspecific functions associated with the reception of speech. Moreover, results in Table 3-III show that both temporal acuity and the TRT are independent of age. In addition, they are mutually independent. Finally, the amount of variance explained by the TRT, expressed in absolute terms, is similar for the normal-hearing and the hearing-impaired group (see Discussion, Section A). This indicates an independent utilization of auditory and non-auditory cues, in accordance with results by Van Rooij and Plomp (1992). Nevertheless, further research should clarify whether the above assumption is valid.

Finally, the Text Reception Threshold should be more widely validated for normal-hearing listeners to determine the sources responsible for the variance in TRT. It is now assumed to be a general measure of non-auditory factors relevant for speech reception, presumably related to a modality-independent skill to perform perceptual closure. Recently, however, a study has started in our laboratory investigating the relation between the SRT and several variants of the TRT for listeners with Dutch as a second language, who are clearly less experienced in applying semantic, syntactic or lexical information to improve the readability or recognition of Dutch sentences. Future studies like these may enable specification of the TRT-measure into less general cognitive or linguistic skills which contribute to speech perception.

V. CONCLUSIONS

The main results arrived at in this study can be summarized as follows:

I. The two groups of normal-hearing and hearing-impaired participants, matched for age, do not perform significantly differently on a non-auditory test (TRT). Differences in speech reception between both groups are thus likely to be mainly governed by auditory factors.

II. Differences in speech reception (SRT) between the normal-hearing and hearing impaired participants are more prominent in modulated noise than in stationary noise. Therefore, modulated noise may be preferred over stationary noise to measure speech reception for clinical purposes, i.e. to assess whether speech recognition in daily life is deteriorated compared to normal.

III. For the normal-hearing listeners, non-auditory factors, as measured by the Text Reception Threshold (TRT), are the most importantsource of variance in the SRT, both in stationary and in modulated noise.

IV. For the hearing-impaired participants, inter-individual differences in speech reception in stationary noise are mainly governed by auditory factors, in particular by the audiogram. In contrast, both auditory (temporal resolution) and non-auditory (TRT) factors account for inter-subject variance in speech reception in modulated noise.

V. The combined measurement of the SRT in modulated noise and the TRT may be clinically relevant to determine part of the origin (auditory or non-auditory) of deteriorated speech reception, possibly adapting the expectations from auditory rehabilitation.

VI. ACKNOWLEDGEMENTS

This research was supported by the Heinsius-Houbolt Foundation, The Netherlands. Thanks are due to Hans van Beek for technical support and for his programming work on the TRT. Nienke Versteeg and Anushka Lala are thanked for their contributions to this experiment.

Clinical relevance of auditory and nonauditory factors affecting speech reception in noise

4

Erwin L.J. George, Sophia E. Kramer, Joost M. Festen and Tammo Houtgast

Submitted for publication in Ear and Hearing

Objective: This study aims at: i) investigating the relevance of auditory and nonauditory factors for the reception of speech in noise, and ii) identifying clinically relevant tests, that account for an as large as possible amount of variance in speech-in-noise reception. *Design:* A battery of tests was constructed, including tests of both auditory functions and nonauditory, verbal and cognitive abilities. Three measures of sentence intelligibility were assessed: the speech reception threshold (SRT) in stationary background noise, the SRT in nonstationary noise and the difference between the two, defined as masking release. Age-matched normal-hearing and sensorineural hearing-impaired, older listeners were selected as participants. *Results:* A principal component analysis showed that two components, one auditory and one nonauditory in nature, could account for the major part of the variance of the test outcomes. Speech reception in noise was found to be mainly governed by the auditory component, but when it comes to speech comprehension in nonstationary noise or to masking release (i.e. listening in the gaps), the nonauditory component also played a significant role. *Conclusions:* For clinical practice, the relevance of auditory and nonauditory factors for speech reception was adequately estimated by the audiogram, a test of spatial working memory (SWM), and the Text Reception Threshold (TRT), a visual analogue of the SRT. The TRT was still significantly associated with speech reception (masking release) after age and audiogram had been partialled out. These findings make it possible to identify the sources responsible for deteriorated speech reception in clinical practice.

I. INTRODUCTION

In clinical audiology, one of the most common complaints of patients visiting an audiological center is that their 'hearing is reasonable, but understanding a conversation in daily life is very difficult'. Clinical investigation often confirms that listening to speech in quiet conditions is fairly easy for these patients, but when a background noise is introduced, speech reception sometimes dramatically deteriorates. Unfortunately, everyday listening situations are seldom very quiet, hence the experienced hearing problem. Thus, the real problem for mild to moderate hearing-impaired patients is often not the hearing loss per se, which can be solved by hearing aid amplification, but the accompanying difficulties with understanding speech in background noise (Plomp, 1978; Duquesnoy and Plomp, 1980; Eisenberg et al., 1995; Divenyi and Haupt, 1997; Bacon et al., 1998; Summers and Molis, 2004). When masker levels fluctuate over time, the difference between normal and hearing-impaired listeners is even more prominent. Apparently, hearing-impaired listeners benefit less from 'listening in the gaps' (Festen and Plomp, 1990; George et al., 2006, 2007). Understanding the processes behind these findings is important, because fluctuating backgrounds are very common in everyday situations (Kramer et al., 1996).

Plomp (1978) formulated a model description for the speech reception threshold (SRT) for hearing-impaired listeners. He distinguished two auditory components: i) hearing loss due to attenuation, related to the hearing threshold and 'solvable' by wearing a hearing aid; and ii) hearing loss due to distortion of the speech, which cannot be solved by simply amplifying the signal. The latter component is considered to be related to so-called suprathreshold auditory deficits, like reduced spectral and temporal resolution, or a loss of auditory compression, which are all regarded to adversely affect speech reception in noise.

However, besides the role of the peripheral auditory organ, perception of speech also involves information processing in the central auditory pathway and cognitive functions, like working memory capacity and speed of information processing (Jerger et al., 1991; Gordon-Salant and Fitzgibbons, 1997; Gatehouse et al., 2003; Lunner, 2003; Hällgren, 2005; Humes, 2002, 2005). Divenyi et al. (2005) indeed show that the deterioration of speech reception with age is accelerated significantly relative to the decline in audiometric measures, suggesting an additional decline of central auditory processing. Moreover, speech reception is also influenced by other nonauditory factors, like an individual's language proficiency (Van Wijngaarden, 2002, 2004) or his familiarity with the sentence context (Warren, 1970; Boothroyd and Nittrouer, 1988; Dubno et al., 2000). Thus, speech recognition is affected by an interaction between, on the one hand, bottom-up or 'stimulus-driven' processes, and, on the other hand, top-down or 'knowledge-driven' factors (Goldstein, 2002). The relative contribution of purely auditory and not purely auditory (cognitive and verbal) processes to speech recognition is, however, still under discussion.

One approach to investigate the contribution of nonauditory processes is to measure speech reception in increasingly complex listening situations, as performed by Pichora-Fuller et al. (1995), who showed that audiometric thresholds alone cannot fully account for the difficulty that the elderly experience in understanding speech in complex environments. They introduced a processing model, in which auditory difficulties adversely affect speech understanding both directly, by altering the amount of correctly perceived words, and indirectly, because effortful listening consumes resources that could otherwise be allocated to nonauditory processes necessary for speech understanding. The relative importance of top-down or cognitive functions seems to increase when speech information is degraded, either by hearing-impairment, or by the presence of background noise or reverberation (Hällgren, 2005; Zekveld et al., 2007a). Nevertheless, audiometric hearing thresholds still seems to be responsible for the major part of the interindividual variance in speech reception, as concluded by for instance Van Rooij and Plomp (1992).

A disadvantage of assessing nonauditory processes with auditory stimuli is that hearing-impairment is a confounding factor in the results. Therefore, an alternative approach to confirm the importance of cognitive factors for speech reception may be followed, in which the relation between auditory performance and visually measured variables is assessed, as performed by for instance Watson et al. (1996) and by Zekveld et al. (2007b). They found significant correlations between overall auditory and nonauditory performance, and suggested that this association likely reflects the shared relevance of a specific cognitive function in both visual and auditory language comprehension. Grant et al. (1998) referred to this function as ‘perceptual closure’, that is, the ability to form linguistic wholes from perceived fragments.

In addition to the results from literature mentioned above, results from an earlier experiment in our laboratory (George et al., 2006) also indicated that nonauditory factors may be related to speech intelligibility in noise. When measuring speech reception thresholds (SRT) for sentences in stationary noise and in several amplitude-modulated maskers, age was found to be one of the main factors (besides temporal resolution) governing masking release for speech in modulated noise, consistent with earlier results by for example Gustafsson and Arlinger (1993) and Snell et al. (2002). It was suggested that this contribution of age to the interindividual variance in 'listening in the gaps' may be associated with cognitive or other nonauditory factors. In a follow-up experiment (George et al., 2007), a visually presented verbal test was used (as in Zekveld et al., 2007b), showing that nonauditory factors are indeed important for the comprehension of sentences in stationary and nonstationary background noise.

The current investigation is an extension to our previous experiments. Besides measures of speech reception, tests of both verbal and cognitive nonauditory functions were performed in a larger group of normal-hearing and hearing-impaired older participants. The current paper's specific goals are: i) to investigate the relevance of auditory and nonauditory factors for the reception of speech in noise and, in particular, masking release, considering the earlier obtained age effect (George et al., 2006), and ii) to identify clinically relevant tests that account for an as large as possible amount of interindividual variance in speech reception and masking release.

To achieve these goals, a battery of tests was constructed, consisting of tests of both auditory functions and nonauditory, verbal and cognitive abilities. Three measures of speech reception were assessed: the SRT in stationary background noise, the SRT in nonstationary noise and masking release, defined as the difference between the two SRTs. Normal-hearing and hearing-impaired participants were matched according to age to increase homogeneity of the total group, thus preventing a possible 'artificial' effect of age on speech reception due to the confounding of age and hearing-impairment. After evaluating differences between the normal-hearing and hearing-impaired subgroups, a Principal Component Analysis was conducted to get a clearer view on the interdependency between the test outcomes. Finally, multiple regression analyses were applied to determine the variables that most adequately account for interindividual differences in speech reception.

II. MATERIALS AND METHODS

A. Participants

Thirty normal-hearing (NH) and thirty sensorineural hearing-impaired (HI) listeners participated in this experiment. Hearing-impaired participants (18 females, 12 males) were patients of the audiology department of the VU University Medical Center,

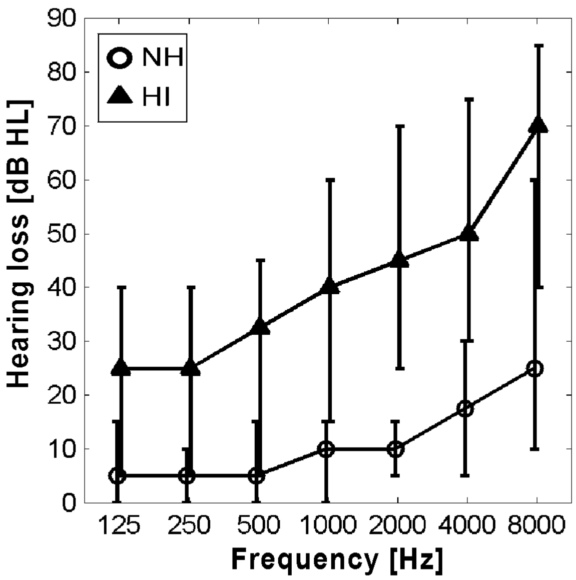

Figure 4-1 Median pure-tone hearing thresholds (*re :* ISO-389:1991) and 5th and 95th percentiles for normal-hearing (NH, circles) and hearing-impaired (HI, triangles) participants.

selected to have a sensorineural hearing loss with average pure-tone thresholds up to 60 dB HL. The age of the hearing-impaired listeners ranged from 46 to 83 years, with an average of 66.2 years. Normal hearing listeners (17 females, 13 males) were acquaintances of the hearing-impaired participants, selected to have pure-tone hearing thresholds better than 15 dB HL at 0.25, 0.5, 1.0 and 2.0 kHz and better than 30 dB at 4.0 kHz. The age of the normal-hearing listeners ranged from 48 to 78 years, with an average of 61.2 years. Figure 4-1 shows the median and the range of the hearing loss for both groups of participants. All participants were native Dutch speakers and reported normal or corrected-to-normal vision. Their color vision was screened with Ishihara plates (Ishihara, 1989).

B. Description of the tests

1. Pure-tone thresholds

Each experimental run started with the measurement of the listener's pure-tone hearing thresholds at octave frequencies between 125 and 8000 Hz, using the same apparatus as during all other measurements. The audiogram was used later as input to shape the spectrum of the auditory stimuli. Only the Pure-Tone Average (PTA) will be included in the data analysis, defined as the average pure-tone hearing threshold at octave frequencies 0.5, 1.0 and 2.0 kHz.

2. Spectral and temporal acuities (F / T)

Each listener's spectral and temporal acuities were determined by making use of an adaptive measurement procedure as introduced and validated by Hilkhuysen et al. (2005). In this procedure, listeners were asked to report the number of tone sweeps (zero to three) they were able to detect in spectral or temporal noise grids. The observed masking release in these noise grids as compared to the threshold in unmodulated noise was used to estimate auditory-filter and time-window widths. In our experiment, both temporal and spectral acuity were determined halfway the listener's dynamic range, in the frequency region around 1 kHz.

Spectral and temporal acuity will be denoted by 'ΔF' and 'ΔT', the capital Greek Delta's emphasizing that threshold-related effects have been accounted for and that the outcome measures are considered to be related to suprathreshold processing in the inner ear. Details on these procedures can be found in our earlier papers (George et al., 2006, 2007).

3. Speech Reception Threshold (SRT)

SRT-measurements were performed using a simple adaptive one up one down procedure for a list of thirteen everyday Dutch sentences (Plomp and Mimpen, 1979) in stationary background noise, as well as in a 16-Hz block-modulated masker. Sentences were read by a female speaker and were unknown to the listener. The masker level was kept fixed, while the speech level was varied adaptively to estimate the SRT, defined as the estimate of the speech-to-noise-ratio at which 50% of the sentences can be reproduced without error. In each condition, the first sentence was presented at a level below threshold and repeated, at 4-dB higher levels with each repetition, until the listener was able to reproduce it correctly. The remaining twelve sentences were presented only once, following an adaptive procedure, with a 2-dB step size. The SRT was estimated as the average signal-to-noise level of sentences number 5 to 14. The fourteenth sentence was not presented, but its level can be calculated from the response to the thirteenth sentence.

To optimize audibility, all auditory signals were spectrally shaped to reach octave masker levels equal to the middle of the dynamic range for each listener. The lower limit of the dynamic range was chosen to be the individual pure-tone threshold, while the upper limit was chosen to be 115 dB SPL for all listeners. The masker level was always the mean of these two. The long-term speech spectrum was shaped accordingly. The bandwidth of noise and speech signals was restricted to frequencies between 223 and 4490 Hz.

SRT in stationary noise and SRT in modulated noise will be used as dependent variables in the data analysis, and are denoted by SRT_{STAT} and SRT_{MOD}, respectively. An additional dependent variable will be masking release, denoted by MR and defined as the difference between SRT_{MOD} and SRT_{STAT}.

4. Text Reception Threshold (TRT)

To determine whether nonauditory processes could play a role in the recognition of speech, a visual equivalent of the SRT was used: the Text Reception Threshold or TRT. This test was developed by Zekveld et al. (2007b), who used it to investigate the relation between auditory and visual perception in normal-hearing subjects. In a follow-up study, George et al. (2007) obtained a significant association between the TRT and speech reception for hearing-impaired listeners, after taking auditory factors into account.

A list of thirteen everyday Dutch sentences, adopted from Versfeld et al. (2000), was presented visually on a computer screen, sentence by sentence, and was distorted by an adaptively changing masking pattern of vertical bars. The amount of masking was varied adaptively, following a one-up-one-down procedure similar to the SRT test, to determine the Text Reception Threshold or TRT, defined as the amount of unmasked text needed by the subject to comprehend 50% of the sentences correctly. In each condition, the first sentence was presented with a percentage of unmasked text below threshold and repeated, adding an extra 12% of unmasked text per step, until the subject was able to reproduce it correctly. The remaining twelve sentences were presented only once, following an adaptive procedure with a 6% step-size. The TRT was estimated as the average percentage of unmasked text of sentences 5 to 14. It is regarded as a general measure of modality-aspecific cognitive and linguistic skills contributing to the perception of partially masked sentences. More details about the test can be found in Zekveld et al. (2007b).

5. Visual monitoring tasks (VMT)

Results by Gatehouse et al. (2003) show a significant interaction between hearing-impairment and cognitive abilities, which were assessed using a visual digit-monitoring task and a visual letter-monitoring task. Considering this result, we decided to include slightly modified versions of both tests in our test-battery. The former was originally developed at the University of Iowa (Knutson et al., 1991), while the latter was developed by the MRC Institute of Hearing Research.

In the digit-monitoring task, subjects were instructed to monitor a stream of single digits between one and nine, presented on a computer screen at a rate of one digit per second. They were asked to press the spacebar of the keyboard before the arrival of the next digit, when three consecutive digits formed an 'odd-even-odd'-sequence. The letter-monitoring task was similar in structure, but here participants were asked to press the spacebar when three consecutive letters formed a three-letter word in the Dutch language. To be able to perform the test in Dutch, we translated the original English version, using the same word-inclusion and -exclusion criteria as the original test. Individual performance of hits versus false alarms on each of the tests was characterized by d' (*d prime*; see e.g. Green and Swets, 1974), a common statistical measure of the correct and false responses to target and nontarget sequences. A higher d' indicates better performance.

Because the separate VMTs (digit or letter) appeared to have relatively low test-retest reliabilities ($r_{tt} < 0.80$) and a relatively high mutual correlation ($r = 0.63$), it was decided to calculate only a single d' as a general measure of visual monitoring, which will be denoted as 'VMT'.

6. CANTAB (RVP / SWM)

The Cambridge Neuropsychological Test Automated Battery (CANTAB) is a set of neuropsychological tests, that is administered by a computer with a touch screen. It was originally developed at Cambridge University in 1986 (Robbins et al., 1994) and is now widely used in clinical settings to examine specific components of cognition. Three subtests of the CANTAB were included in our test-battery. They provide a total of four outcome measures (RVP-A, RVP-B, SWM-B and SWM-S), which are described below.

Each CANTAB-trial started with a test of motor screening, which is a training procedure to relax participants and to introduce them to the computer and the touch screen. It simultaneously screened for difficulties with vision, movement or comprehension and ascertained that the subject could follow simple instructions. Subjects were instructed to touch a cross appearing on the screen, after which the cross was replaced by another cross at a different location on the screen. The test finished after ten presentations. No abnormalities were found for any of the subjects.

The Rapid Visual Processing test (RVP) is a test of sustained visual attention. Moreover, it is sensitive to each subject's general cognitive performance. Subjects were requested to detect target sequences of digits from a stream of numbers 2 to 9, appearing on the computer screen in random order at the rate of 100 digits per minute. Target sequences were 2-4-6, 3-5-7 and 4-6-8, which occurred at the rate of 16 every two minutes. Each test run included a total of 27 targets. The RVP test produced two outcome measures, namely A' (*A prime*) and B'' (*B double prime*). A' is a signal detection measure of sensitivity to the target, regardless of response tendency. It ranges from 0 to 1; a score closer to 1 indicates that the subject is better at detecting targets. B'' is a signal detection measure of the strength of trace required to elicit a response (range –1.00 to +1.00). Thus, it is the tendency to respond regardless of whether a target sequence was present. A score close to +1.00 indicates fewer false alarms, thus a better performance. In the data analysis, A' and B'' are denoted by 'RVP-A' and 'RVP-B', respectively.

The Spatial Working Memory test (SWM) is a test of the subject's ability to retain spatial information and to manipulate remembered items in working memory. This test is a sensitive measure of frontal lobe and executive dysfunction and also assesses the use of an efficient strategy. The aim of the test was that the subject should find a blue 'token', that was hidden in one of the boxes displayed on the screen. After a token had been found, a next token was hidden inside one of the boxes not previously used during that trial. Subjects had to search for tokens whilst not returning to boxes where a token had previously been found. The number of boxes was gradually increased, until it was necessary to search a total of eight boxes. The color and position of the boxes used were changed from trial to trial to discourage the use of stereotyped search strategies.

The SWM test produces two outcome measures. The first is defined as the number of 'between errors', that is, the number of times that the subject revisited a box in which a token had previously been found. The second outcome measure denotes the extent to which a subject used an efficient search strategy. Its calculation is based on research by Owen et al. (1999), who suggested that an efficient strategy for completing this task is to follow a predetermined search sequence. The strategy score is estimated by counting the number of times a subject starts a new search in a different box. A high score (maximum is 56) represents a poor strategy score, while a low score (minimum is 8) equates to an effective use of strategy. The between errors score and the strategy score are denoted by 'SWM-B' and 'SWM-S', respectively.

C. Instrumentation and general procedure

The experiment was run on a Dell personal computer, equipped with an Elo 16-inch touch-sensitive display, a Creative Labs Audigy external sound device and Beyer Dynamic DT48 headphones. Sound calibrations were performed with a Brüel & Kjær Artificial Ear (type 4152) and a Brüel & Kjær 2260 Observer conform ISO 389 (1991). All measurements were performed while listener and investigator were seated in a sound-insulated room.

All measurements, except the CANTAB, were performed following a test-retest design in a single three-hour session, interrupted by several small breaks. Test and retest blocks each included the measurement of VMT (digit), VMT (letter), three TRT, two SRT_{MOD}, SRT_{STAT}, F and T. Measurements of the TRT and the SRT_{MOD} were performed more than once in both test and retest to improve reliability. Test and retest outcomes were averaged. CANTAB-measurements were performed only once, between the test and retest blocks; evaluation of their reliabilities was considered unnecessary in the light of their extensive use in other research areas. The order in which the tests were presented was fixed. A session always started with the measurement of the audiogram and visual screening. Auditory measurements were conducted monaurally, using the participant's best ear, which was chosen according to his or her audiogram, or, in case of doubt, personal preference in telephone conversation.

D. Statistical analysis

All analyses were performed using SPSS for Windows, release 11.0.1. Before conducting any analyses, the data were screened for test-retest inconsistencies and severe outliers were removed. Variables were transformed (by taking the logarithm or square root) in cases where normality was improved by the transformation. Two-tailed *t*-tests were conducted to evaluate differences in test-outcomes between normal-hearing and hearing-impaired participants.

Correlations between the test outcomes were investigated. Because it is common practice to adopt a strict criterion for the significances when performing multiple comparisons or correlations (Miller, 1981), only effects reaching *p*-values below 0.01 were considered significant.

A Principal Component Analysis (PCA) was performed with Varimax rotation and Kaiser normalization, based on only those independent variables that were significantly related to speech reception. The results of the PCA were used to get a clearer view on the interdependency between the variables, and to investigate the relation between speech reception and the derived components. Finally, multiple stepwise regression analyses (MSRs) were performed to identify tests that best account for interindividual differences in speech reception.

The PCA and the regression analyses were conducted for all data, and not for normal-hearing and hearing-impaired participants separately, for two reasons. First, the number of subjects in the current experiment (2x30) was rather small compared to the number of variables; pooling all data gives rise to a better subjects-to-variables ratio (see e.g. Grimm and Yarnold, 1995). Secondly, the results show that there is an overlap in test scores between the two subgroups. Instead of representing two clearly distinguishable groups, most test scores appear to be forming a continuum. Therefore, results of the two subgroups will not be dealt with separately, although, for clarity, the distinction is preserved in displaying the results.

III. RESULTS

A. Overview of test outcomes

Table 4-I shows descriptives of the test outcomes, ordered by origin (auditory or nonauditory). These data were screened for test-retest inconsistencies and severe outliers. When, for a specific participant, either the difference between test and retest, or the average score, was more than three standard deviations away from the group average, then that specific score was excluded from further analyses. Transforming the data, as described below, did not help in reducing the number of outliers.

Subsequently, the normality of the variables SRT_{STAT}, SRT_{MOD}, MR, ΔT and PTA was improved by taking their natural logarithm, while SWM-B was more normally distributed when it's square root was taken. Therefore, all further analyses were conducted using the transformed values of these test outcomes.

To compare the normal-hearing and hearing-impaired participants, Table 4-II displays an overview from the test outcomes for the two subgroups, plus the results from the *t*-tests. The *t*-statistics show that normal-hearing and hearing-impaired participants perform significantly differently on auditory tests, like temporal acuity and audiogram (PTA). The latter is not surprising because both subgroups were selected on the basis of their differences in audiometric thresholds. Spectral acuity (ΔF), however, does not seem to differentiate between normal-hearing and hearing-impaired listeners, as shown before by George et al. (2006). Moreover, the normal-

		unit	ALL N	M	S	r_{tt}	SEM
Auditory	SRT_{STAT}	dB SNR	58	-3.1	2.0	.86	0.7
	SRT_{MOD}	dB SNR	59	-12.9	5.1	.96	1.0
	MR	dB	58	-9.8	3.3	.83	1.4
	ΔF	Hz	57	201.7	46.9	.85	17.9
	ΔT	ms	57	5.43	3.02	.97	0.49
	PTA	dB HL	60	16.3	23.1	-	-
Nonauditory	age	years	60	63.7	10.1	-	-
	TRT	% text	60	59.5	3.3	.86	1.2
	VMT	-	60	2.02	1.95	.90	0.63
	RVP-A	-	60	0.89	0.05	-	-
	RVP-B	-	59	0.93	0.06	-	-
	SWM-B	# errors	60	29.9	22.6	-	-
	SWM-S	-	60	33.6	6.2	-	-

Table 4-I Averages (M), standard deviations (S), test-retest reliabilities (r_{tt}), standard errors of measurement (SEM) and total number of valid observations (N) of the (transformed) test outcomes, as calculated for all participants. Test reliabilities r_{tt} have been calculated from test-retest correlations r_{tr} using the Spearman-Brown formula: $r_{tt} = 2 * r_{tr} / (1 + r_{tr})$, cf. Nunnally, 1967. SEM is defined as $S * \sqrt{(1 - r_{tt})}$.

hearing and hearing-impaired participants do not significantly differ in age, which is a direct consequence of the age-matching. Finally, the *t*-statistics show that the subgroups do not perform significantly different on the nonauditory tests.

The fact that differences between the subgroups are mainly auditory in nature indicates that interindividual differences in speech reception are also likely to be mainly governed by auditory factors. The finding that the subgroups' age and nonauditory performance do not differ significantly is reassuring. It means that the matching of age has been successful: if a significant relation between nonauditory performance and speech reception can now be identified, this association can be

			NH					HI						
		unit	N	M	S	r_{tt}	SEM	N	M	S	r_{tt}	SEM	*t*	*p*
Auditory	SRT_{STAT}	dB SNR	30	-4.0	1.2	.60	0.8	28	-2.0	2.3	.86	0.9	5.135	< .001
	SRT_{MOD}	dB SNR	30	-15.1	2.2	.86	0.8	29	-9.7	5.3	.95	1.2	7.223	< .001
	MR	dB	30	-11.1	1.7	.54	1.1	28	-8.0	4.0	.78	1.9	5.574	< .001
	ΔF	Hz	30	195.7	30.2	.85	11.9	27	208.4	60.2	.86	22.9	0.990	.329
	ΔT	ms	30	4.39	1.87	.96	0.37	27	6.88	3.51	.97	0.62	4.397	< .001
	PTA	dB HL	30	7.7	5.0	-	-	30	34.4	17.6	-	-	12.646	< .001
Nonauditory	age	years	30	61.2	9.6	-	-	30	66.2	10.1	-	-	1.970	.054
	TRT	% text	30	59.2	3.3	.85	1.3	30	59.7	3.2	.88	1.1	0.545	.588
	VMT	-	30	2.32	1.10	.90	0.36	30	1.73	0.86	.88	0.29	2.303	.025
	RVP-A	-	30	0.90	0.06	-	-	30	0.89	0.05	-	-	0.777	.440
	RVP-B	-	29	0.93	0.06	-	-	30	0.94	0.06	-	-	0.769	.444
	SWM-B	# errors	30	25.7	22.5	-	-	30	34.4	21.5	-	-	1.763	.083
	SWM-S	-	30	33.4	6.4	-	-	30	33.8	6.0	-	-	0.229	.820

Table 4-II As Table 4-I, but calculated for the normal-hearing (NH) and the hearing-impaired (HI) participants separately. Differences between the subgroups, as expressed by the *t*-statistics, were considered significant if $p < 0.01$.

considered to be a result of interindividual differences and not of a confounding effect of subgroup.

B. Correlations between test outcomes

To investigate the interrelationships between the variables, a correlational analysis was performed. Before the analysis, the scores of VMT, RVP-A and RVP-B were inverted, such that lower values represent a better score for all variables. A significant positive signed correlation between two variables then indicates that a positive performance on one test accompanies a positive performance on the other test.

		Auditory						Nonauditory						
		SRT_{STAT}	SRT_{MOD}	MR	ΔF	ΔT	PTA	age	TRT	VMT	RVP-A	RVP-B	SWM-B	SWM-S
Auditory	SRT_{STAT}													
	SRT_{MOD}	$\underline{\underline{.78}}$												
	MR	$\underline{\underline{.44}}$*	$\underline{\underline{.71}}$*											
	ΔF	.04	.10	.12										
	ΔT	$\underline{.37}$	$\underline{\underline{.57}}$	$\underline{\underline{.57}}$	-.30									
	PTA	$\underline{\underline{.59}}$	$\underline{\underline{.73}}$	$\underline{\underline{.66}}$	-.04	$\underline{\underline{.65}}$								
Nonauditory	age	$\underline{.40}$	$\underline{\underline{.42}}$	.32	.03	.15	.31							
	TRT	.23	$\underline{.38}$	$\underline{\underline{.43}}$	-.09	.26	.25	.27						
	VMT	.22	$\underline{.34}$	.33	.07	.03	.30	$\underline{\underline{.44}}$	$\underline{.35}$					
	RVP-A	.07	.21	.30	-.02	.01	.21	$\underline{.35}$	$\underline{.37}$	$\underline{\underline{.64}}$				
	RVP-B	-.18	-.11	-.01	-.15	-.01	-.01	.23	$\underline{.34}$	$\underline{\underline{.34}}$	--*			
	SWM-B	$\underline{.38}$	$\underline{\underline{.43}}$	$\underline{.34}$	.15	.02	.25	$\underline{\underline{.60}}$	.27	$\underline{\underline{.62}}$	$\underline{\underline{.53}}$	.18		
	SWM-S	.27	.22	.10	.09	-.06	.06	$\underline{\underline{.59}}$	.11	$\underline{.38}$	$\underline{.35}$	.16	--*	

* *Because MR is defined as the difference between the two SRTs, the correlations between the SRTs and MR might have been distorted by shared error variance. Therefore, they were calculated based on the test-values of the SRTs and the retest-values of MR, and vice-versa. This means that the correlations are 'real', and not artificially high. For the same reason, correlations between SWM-B and SWM-S, and RVP-A and RVP-B were not determined.*

Table 4-III Product-moment cross-correlations between the test outcomes. Underlined values are significant at the one percent level, while double underlined values indicate the $p < 0.001$ level.

Correlations are displayed in Table 4-III. Correlation coefficients greater than 0.34 are significant at the 1-percent level, while coefficients exceeding 0.42 are significant at the $p < 0.001$ level. Because speech reception is our main interest, three variables are regarded as 'dependent', namely speech reception in stationary noise (SRT_{STAT}), speech reception in modulated noise (SRT_{MOD}) and masking release (MR), the correlations with which are displayed in the leftmost three columns of Table 4-III. All other variables are regarded as 'independent' variables.

First, not surprisingly, it can be seen in Table 4-III that the measures of speech reception themselves are interrelated. In particular, the correlations of SRT_{STAT} and

MR with SRT_{MOD} are high ($r > 0.70$), while the correlation between SRT_{STAT} and MR is somewhat more modest ($r = 0.44$), although still significant.

Secondly, Table 4-III shows that auditory factors (PTA and ΔT) are significantly related to these measures of speech reception ($r > 0.50$). Nevertheless, nonauditory factors also play a role in the reception of speech, considering the significant correlations between, for instance, SWM-B and SRT_{MOD} ($r = 0.43$) and TRT and MR ($r = 0.43$). It should be noted, however, that the correlations between test outcomes and speech reception may be affected by the mutual interrelationships between the test outcomes, and thus, may be artificially high or 'induced'. For instance, the effects of temporal acuity (ΔT) and audiogram (PTA) cannot be fully distinguished because these two test outcomes are significantly related to each other ($r = 0.65$). Similarly, many nonauditory test outcomes are interrelated, as indicated by the underlined values in the lower right quadrant of Table 4-III.

When considering only the nonspeech variables, significant correlations only appear at the upper left and the lower right quadrant in Table 4-III. This means that the auditory variables are related to each other, and the nonauditory variables are mutually dependent, but there are no significant correlations between the auditory and the nonauditory independent variables. This indicates that it might be possible to split up the variance of the test outcomes in two independent components, one related to auditory processing and the other related to nonauditory processing. This indication will be further investigated in the next section.

C. Principal component analysis

In general, a principal components analysis (PCA) can be used to visualize the relations between various test scores and to reduce the number of variables. Theoretically, the ten measured 'independent' scores for each participant can be represented by a single point in a ten-dimensional space, with each independent variable plotted along its own dimensional axis. For sixty subjects, this results in a cloud of sixty points in a ten-dimensional space. When the variables are mutually independent, each test outcome will account for 10% of the total variance. When the test outcomes are mutually related, however, it is possible to find a new axis that accounts for a larger part of the interindividual variance. Thus, a first principal component is determined by finding the direction in space in which the scores are most widely separated. The first component is then a linear combination of observed variables that explains a maximum amount of the total variance in the scores. The second principal component is orthogonal to the first and chosen to account for as much as possible of the remaining variance. Subsequent components can be chosen in a similar way.

The PCA produces a vector for each derived component, which gives the coordinates of the component's direction in terms of the original variables. The squared length of the vector (or 'eigenvalue') indicates the amount of variance in that direction, while the coordinates of the vector (or 'loadings') display the relative importance of each test score for the component. The loadings can also be interpreted as correlations between the original test scores and the components. Often, most of the variance in scores can be explained by only a small number of components. In our case, only components with eigenvalues larger than unity are considered relevant, that is, components that account for a larger than proportional part of the total variance.

A PCA was performed using the scores for the ten independent variables as inputs. Because our main interest was the relation with speech reception (SRT_{STAT}, SRT_{MOD} and MR), this PCA was based only on the scores of the variables that were significantly related to any of the three dependent variables, as shown in Table 4-III.

	Loading on component #	
	1	2
VMT	.845	.075
SWM-B	.843	.045
RVP-A	.800	.017
age	.677	.227
TRT	.463	.374
ΔT	-.066	.919
PTA	.240	.840
% variance accounted for	40.0	25.0
Cum. variance accounted for	40.0	65.0

Table 4-IV Results of the PCA, based only on the seven 'independent' variables that are relevant for the reception of speech in noise. Displayed are test score loadings on each of the two principal components after Varimax rotation, plus the percentage of variance accounted for. Variables are ordered by loading size, not by origin. Underlined values are significant at the one percent level, while double underlined values indicate the $p < 0.001$ level.

This means that the variables ΔF, RVP-B and SWM-S were excluded from the analysis. RVP-A was still contained because it's relation with MR approaches significance (r = 0.30, p = 0.02). Two principal components were derived, which account for 40% and 25% of the total variance of the test scores, respectively, which sum to a total of 65% explained variance.

Table 4-IV displays the loadings of the included variables on the two derived components. Nonauditory tests are significantly related to principal component 1, while component 2 is mainly related to auditory tests. Thus, component 1 represents variance of the nonauditory variables, while component 2 is mainly auditory in nature. The TRT is the only variable which significantly loads on both the principal components.

Figure 4-2 shows the individual normalized scores on the derived auditory and nonauditory principal components for both normal-hearing and hearing-impaired listeners. The ranges for both subgroups are similar on the nonauditory component, but differ on the auditory component. This finding is in line with the results from the t-statistics, as reported in Table 4-II.

In Figure 4-3, the two derived principal components are used to visualize the relations between all measured variables. Each variable is represented by a single point, while its position in the figure shows the relative proportion of its variance

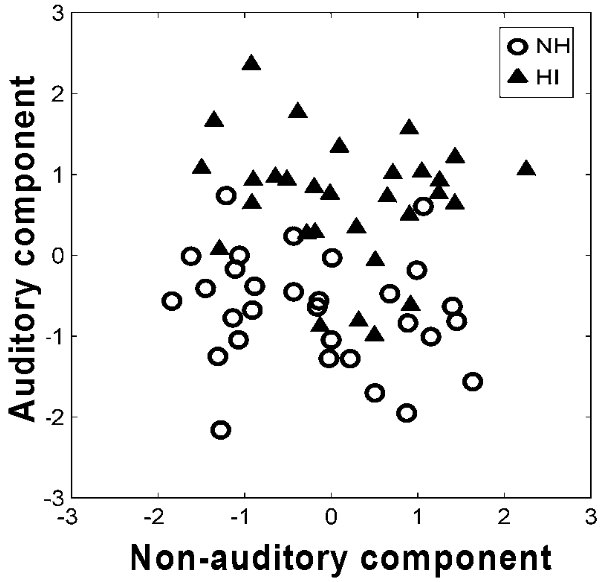

Figure 4-2 Individual normalized scores for the derived auditory and nonauditory principal components, for normal-hearing (NH, circles) and hearing-impaired (HI, triangles) participants.

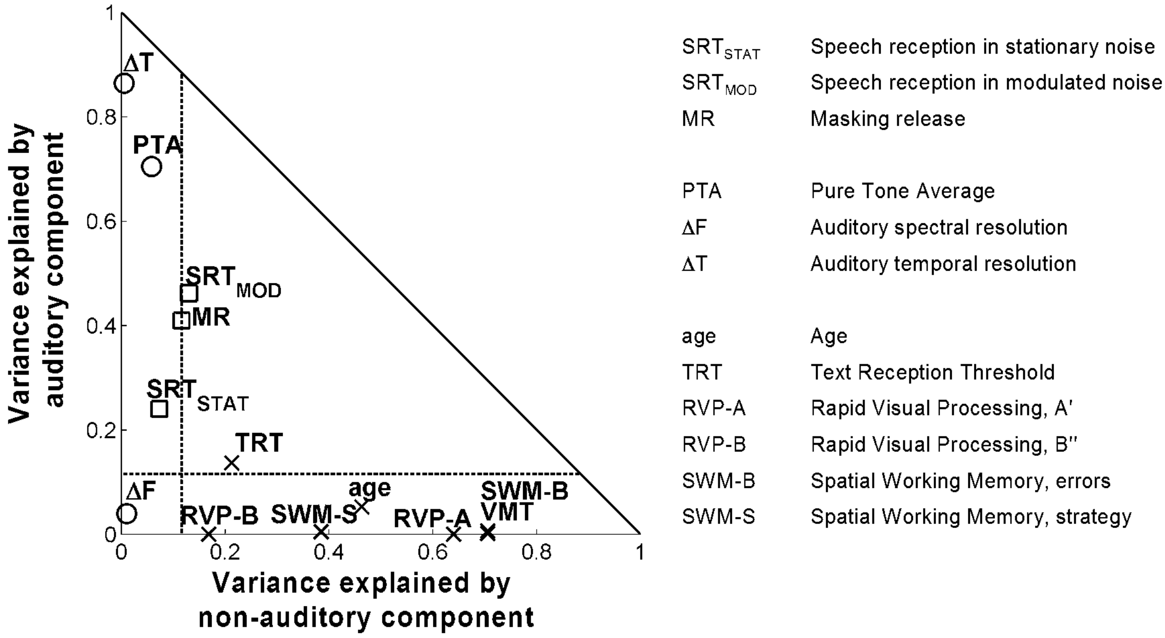

Figure 4-3 Relative proportions of variances (r^2) in test scores, that are accounted for by the two principal components, derived by a PCA based on only those variables that were relevant for speech reception (see text). Auditory test scores are represented by circles, visually measured test scores by crosses, and scores related to speech reception by squares. Dashed lines indicate where the amount of variance, explained by one of the components, reaches significance at the one-percent level, that is, where $r = 0.34$ / $r^2 = 0.12$.

explained by the derived auditory (vertical) and nonauditory (horizontal) components. The distance between a specific point and the diagonal (the x+y=1 line) indicates the amount of variance left unexplained by the two principal components. As expected, variables that are nonauditory in nature (e.g. age, SWM-B and VMT) are positioned close to the horizontal axis, while auditory test scores (like PTA or ΔT) are clustered along the vertical axis. This figure can be regarded as a confirmation of our suggestion that nonauditory and auditory scores are independent (as expressed by the correlations in Table 4-III). As mentioned before, the TRT is the only exception to this 'rule of thumb', loading significantly on both components. This finding is regarded as a confirmation of our earlier suggestion that the TRT reflects a modality-independent perceptual closure skill.

D. Principal components and speech reception

Figure 4-3 not only visualizes the mutual relations of the various variables, but also the relation between the ten independent variables and the three speech-related, dependent variables. All three outcomes of speech reception (SRT_{STAT}, SRT_{MOD} and

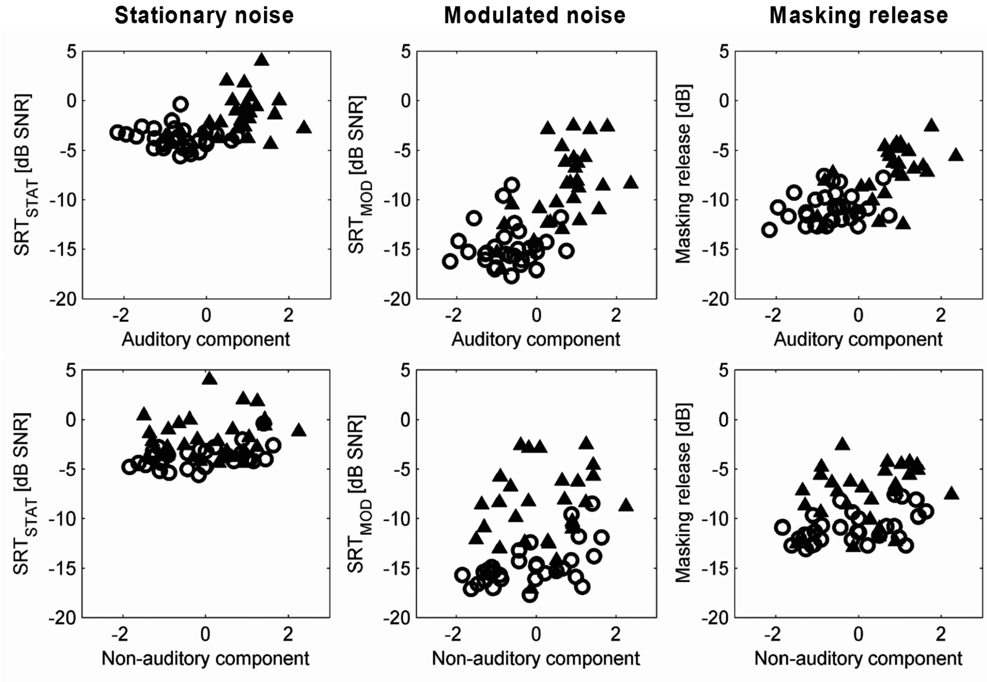

Figure 4-4 Speech reception in stationary noise (SRT_{STAT}), speech reception in modulated noise (SRT_{MOD}) and the difference between the two (masking release), vs. the derived auditory and nonauditory principal components, for normal-hearing (circles) and hearing-impaired (triangles) participants.

		SRT_{STAT}		SRT_{MOD}		MR	
		predictor	R^2	predictor	R^2	predictor	R^2
Components	step 1	AUD	.242	AUD	.448	AUD	.397
	step 2	-	-	NAUD	.575	NAUD	.510
Test outcomes	step 1	PTA	.310	PTA	.505	PTA	.405
	step 2	-	-	SWM-B	.562	TRT	.472

Table 4-V Successive contributors to explaining the variance in SRT in stationary noise (SRT_{STAT}), SRT in modulated noise (SRT_{MOD}) and masking release (MR). Shown is the amount of variance, corrected for the available degrees of freedom, that the derived components or the relevant variables cumulatively account for when included in the model. All shown models have a significance $p < 0.001$. AUD - auditory component; NAUD - nonauditory component.

MR) are significantly related to the auditory component, but a significant part of the variance in speech reception in modulated noise and in masking release is also associated with nonauditory factors.

Figure 4-4 shows the scatter plots of the relations between the three speech outcomes and the derived auditory and nonauditory components. As displayed in Figure 4-3, the auditory component accounts for 24.2% of the variance in speech reception in stationary noise, while nonauditory factors do not contribute here. In modulated noise, the auditory component also explains the major part of the variance (44.8%), but the nonauditory component significantly contributes an extra 12.7%. Because by definition, the two derived components are independent, these two percentages can be summed to give a total of 57.5% of explained variance. The pattern for masking release is similar: 39.7% of the variance is accounted for by the auditory component and the nonauditory component adds an extra 11.3%, summing to a total of 51.0% of explained variance. These results are summarized in the upper part of Table 4-V.

The contribution of auditory and nonauditory factors to SRT_{MOD} and MR is in line with our earlier results (George et al., 2007), which showed that, for normal-hearing listeners, the main part of the variance in speech reception could be attributed to nonauditory factors. In contrast, speech reception within a group of hearing-impaired listeners was shown to be mainly related to auditory factors. Combining these observations, it seems only logical that, when the data from all participants are considered, variance in speech reception is mainly governed by the auditory component, while the nonauditory component also plays a significant, although minor, role.

E. Test outcomes and speech reception

In clinical practice, it is not possible to derive auditory and nonauditory components. Instead, it is necessary to identify and perform individual tests, that can account for an as large as possible part of the interindividual variance in speech reception. Moreover, it is important that the test is efficient, because clinical measurement time is in most cases limited.

Multiple Stepwise Regression Analyses (MSRs) were performed to investigate which individual tests most adequately explain the variance in each of the three measures of speech reception. Results, as displayed in the lower part of Table 4-V, show that PTA accounts for the main part of the explained variance in all three

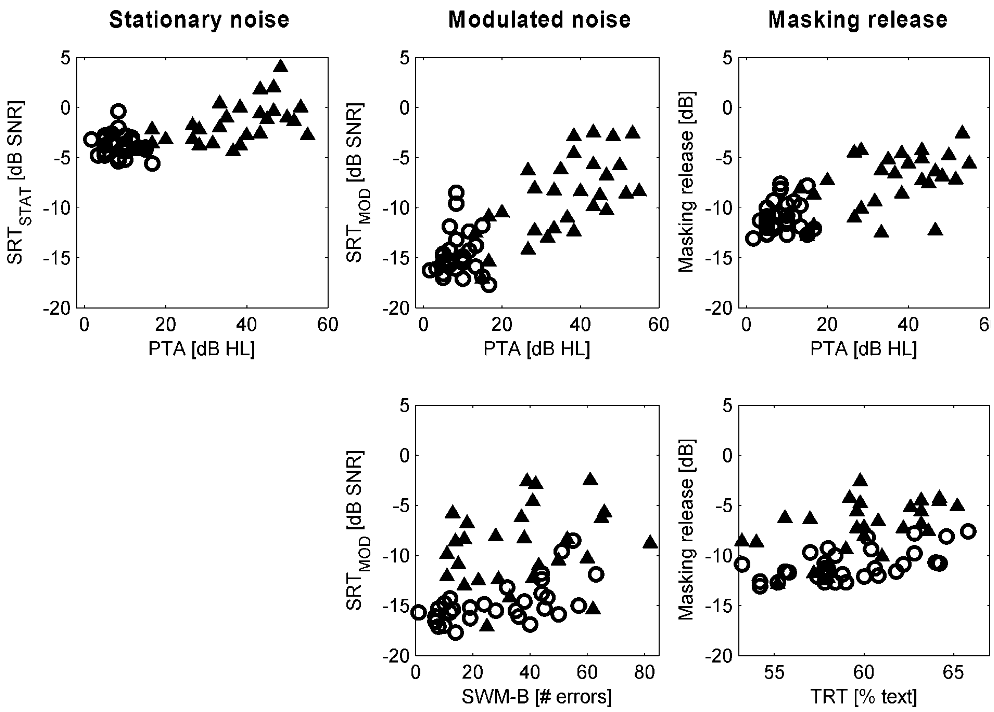

Figure 4-5 Speech reception in stationary noise (SRTSTAT), speech reception in modulated noise (SRTMOD) and the difference between the two (masking release), vs. the variables relevant for speech reception, as derived by multiple stepwise regression analyses (MSRs), for normal-hearing (circles) and hearing-impaired (triangles) participants. In stationary noise, none of the variables significantly contributed to explaining variance after the first step in the MSR, so the corresponding panel is missing.

dependent variables, while SWM-B and TRT account for an additional amount of explained variance in SRTMOD and MR, respectively. After this second step in the MSR, no other independent variables significantly contributed.

The fact that PTA is the best predictor of speech-in-noise reception could be expected, because the derived auditory component, on which PTA heavily loads, explains the largest part of the variance in speech reception. Moreover, the large contribution of PTA is in line with earlier results (see Discussion). Temporal acuity (ΔT) also loads heavily on the auditory component, but PTA and ΔT are related ($r = 0.65$) and the correlations of PTA with the SRTs and MR are simply somewhat higher (Table 4-III). It is convenient that PTA seems the best predictor of speech reception, because measuring the audiogram is already a common first step in clinical evaluation of hearing problems.

To provide somewhat more insight in the choices made in the second step of the MSRs, Table 4-VI displays the correlations between the test outcomes and the three measures of speech reception when the effect of PTA has been partialled out (three leftmost columns). Consistent with the results of the MSRs, no variables significantly add to explaining variance in SRT_{STAT} (stationary noise) when the audiogram has been included in the model. In modulated noise, however, it can be seen that SWM-B is relevant for speech reception in modulated noise (SRT_{MOD}, $r = 0.36$), when the effect of PTA is partialled out, and TRT is significantly associated with masking release (MR, $r = 0.36$).

Figure 4-5 shows the scatter plots of the relations between the identified relevant tests and the three speech outcomes. The resulting amounts of variance in speech reception explained by either the auditory and nonauditory components, or by the identified relevant test outcomes are shown in Table 4-V. It can be seen that the test outcomes account for similar amounts of explained variance as the components. Thus, it can be concluded that, for predicting speech reception in clinical practice, it is unnecessary to fully derive the auditory and nonauditory components. Instead,

		PTA partialled out			PTA & age partialled out		
		SRT_{STAT}	SRT_{MOD}	MR	SRT_{STAT}	SRT_{MOD}	MR
Auditory	ΔF	.08	.19	.20	.07	.19	.19
	ΔT	-.02	.19	.24	.00	.22	.26
	PTA	-	-	-	-	-	-
Nonauditory	age	.28	.30	.17	-	-	-
	TRT	.10	.30	<u>.36</u>	.04	.25	<u>.34</u>
	VMT	.06	.19	.18	-.05	.09	.13
	RVP-A	-.07	.08	.22	-.17	-.01	.18
	RVP-B	-.21	-.14	.00	-.29	-.23	-.04
	SWM-B	.30	<u>.36</u>	.24	.18	.25	.18
	SWM-S	.29	.26	.08	.16	.10	-.02

Table 4-VI Product-moment cross-correlations between the test outcomes and measures of speech reception (SRT_{STAT}, SRT_{MOD} and MR), after partialling out the effect of PTA (left) or the effects of both PTA and age (right). Underlined values are significant at the one percent level.

their relevance for speech reception is adequately estimated by the audiogram (PTA) and the tests of spatial working memory (SWM-B) and perceptual closure in text reception (TRT).

F. Effect of age

In the current experiment, normal-hearing and hearing-impaired participants were age-matched to increase homogeneity of the total group. However, age is still significantly related to both SRT_{STAT} and SRT_{MOD}. This relation is no longer significant when the audiogram (PTA) is partialled out (Table 4-VI), indicating that age is mainly associated with speech reception indirectly, via the deterioration of pure-tone thresholds with age.

Age was also still significantly related to several nonauditory measures, like SWM-B ($r = 0.60$) or VMT ($r = 0.44$), even when the audiogram had been partialled out. This may indicate that the significant relation between SWM-B and SRT_{MOD}, as obtained in Section III.E, might still be influenced by the effect of age. To investigate this suggestion, Table 4-VI also shows the correlations between the test outcomes and the measures of speech reception when, besides PTA, age is also partialled out (three rightmost columns). It can be seen that the correlation coefficient between SWM-B and SRT_{MOD} is no longer significant ($r = 0.25$, $p = 0.06$). This holds for all other independent variables, except for TRT, which is still significantly associated with masking release ($r = 0.34$). This means that a large part of the shared variance in SWM-B and SRT_{MOD} is related to age, while the shared variance in TRT and masking release is not related to age or audiogram effects.

IV. DISCUSSION

A. Comparison with literature

The current results show that PTA is a major contributor to explaining variance in speech reception, in agreement with results by Duquesnoy (1983), Van Rooij and Plomp (1992), and Smoorenburg (1992). Because audibility was optimized in the current experiment, this large effect of PTA is probably not related to a loss of available speech information, that is, not a result of the audiogram per se, as also suggested in our earlier paper (George et al., 2007). Instead, it can be understood by considering PTA as a good estimate for general auditory performance: the development of hearing loss (PTA) generally accompanies deterioration of suprathreshold processing, and vice-versa. As discussed in Section E of the Results, PTA is indeed significantly related to temporal acuity and it is not unlikely that a higher PTA also reflects deteriorated auditory compression, intensity coding, or recruitment (Stephens, 1976).

After taking the effect of hearing loss into account, nonauditory variables explain an additional amount of variance in speech reception, in particular the SWM-B and the TRT tests. A large part of the shared variance in SWM-B and SRT_{MOD} was shown to be related to age. Thus, the ability to manipulate working memory, as measured by the SWM-B, is relevant for speech reception in noise, but its significance might be estimated by age. Only the TRT, a modality-independent measure of perceptual closure, was still significantly associated with speech reception (masking release) after age and audiogram had both been partialled out. These findings can be considered consistent with the results from our earlier experiment (George et al., 2006), where it was suggested that, besides temporal acuity, the obtained significant contribution of age to masking release or 'listening in the gaps' might be associated with cognitive or other nonauditory factors.

Our results do not seem to be in line with results by Gatehouse et al. (2003), who showed that listeners who perform better on the VMT-task derive greater benefit from the temporal structure in background noise, that is, they are better at 'listening in the gaps'. A possible explanation for this dissimilarity is that, in our study, participants were both normal-hearing and hearing-impaired listeners, while Gatehouse and coworkers only used hearing-impaired listeners, who, as a second difference, were established users of hearing-aids. In addition, their observed association between VMT with speech reception may be attributed to the shared variance in VMT and age (see Table 4-III) or to the confounding effect of audiogram. When comparing Table 4-III and 4-VI, it can be seen that VMT is significantly related to SRT_{MOD} at first, but this relation is not significant anymore when the audiogram (PTA) has been partialled out.

B. Relations among measures of speech reception

It was shown is Section III.B that that the three measures of speech reception, as used in the current study, are interrelated. The association between SRT_{STAT} and SRT_{MOD} is significant (Table 4-III, $r = 0.78$) , which can be understood because they are both measures of the ability of a listener to efficiently detect and understand speech in noise. Moreover, SRT_{MOD} and masking release are mutually related ($r = 0.71$), which can be explained because both measures rely on a listener's ability to 'listen in the gaps', that is, to benefit from the relatively silent periods in the modulated masker.

In this light, it is interesting that the correlation between SRT_{STAT} and MR seems smaller (r = 0.44), although significant. In fact, when the effect of audiogram is partialled out, the relation between SRT_{STAT} and MR is nonsignificant (r = 0.19, p = 0.05), while the other two correlations still are (r > .50, p < 0.001). This finding indicates that the processes responsible for speech reception in stationary noise and those responsible for the ability to 'listen in the gaps' are only loosely related, and might be different in nature. Our result that the derived nonauditory component significantly contributes to variance in SRT_{MOD} and MR, but does not contribute to speech in reception in stationary noise, is regarded as a confirmation of this indication.

Thus, to fully understand the difficulties that hearing-impaired listeners experience when listening to speech in noise, measuring speech reception in nonstationary noise is not only important because nonstationary maskers are more representative for everyday listening situations (Kramer et al., 1996), but also for identifying the nonauditory component relevant for speech reception. Moreover, differences between normal-hearing and hearing-impaired listeners are more prominent in nonstationary maskers (Festen and Plomp, 1990; George et al., 2007).

C. Clinical relevance

It has been concluded above that measuring SWM-B or TRT significantly contributes to accounting for variance in speech reception in modulated noise or masking release, respectively. Moreover, SWM-B and TRT are major sources of variance in SRT_{MOD} or masking release within the normal-hearing group, as reconsideration of the lower right panel of Figure 4-5 suggests (r = 0.58 for both cases). The relevance of the TRT is in line with results by Zekveld et al. (2007b) and, for instance, Amitay (2002), who also obtained significant associations between the comprehension of written and spoken language. This shared component likely reflects a modality-independent skill to perform perceptual closure, related to the extent to which listeners use the redundancies in speech at the acoustic, phonetic and lexical level (Warren, 1970).

The fact that SWM-B and TRT vary substantially across listeners may enable applying them for clinical purposes (see also George et al., 2007). As can be seen in Figure 4-5, the SRT_{MOD} or MR score obtained by normal-hearing participants, at a specific SWM-B or TRT-value, might be considered as a reference or a 'limit of best performance' for hearing-impaired listeners. When a datapoint is close to this limit, SRT_{MOD} or MR can be regarded essentially normal (i.e. mainly related to nonauditory factors). An individual's vertical deviation from the normal-hearing reference may then be interpreted as the relative contribution of auditory factors to deteriorated speech reception or masking release. In summary, the SWM-B and the TRT tests do not only contribute significantly to explaining variance of speech-in-noise reception, but also make it possible to determine the relative contribution of auditory and nonauditory factors to speech reception, applying them as an additional clinical tool.

D. Factors affecting the TRT

Results from the current experiment show that the TRT significantly loads on both the auditory and nonauditory components. In addition, it is the only variable that is still significantly associated with speech reception after age and audiogram have been partialled out. The TRT is now assumed to be a general measure of nonauditory factors relevant for masking release, presumably related to a modality-independent skill to perform perceptual closure. In the light of the current results, however, it would be interesting to be able to specify the TRT-measure in terms of more specific cognitive or linguistic skills.

Reconsideration of Table 4-III, which displays the correlations between all variables, shows that the TRT is significantly associated with VMT and the RVP-measures, while correlations with SWM-B and SWM-S are not significant. It seems that the nonauditory cognitive and verbal skills, as measured by the TRT, are more related to the rapid processing of visual information (the common component in the VMT and RVP tests) than to working memory load (as measured by the SWM test). Future studies will have to be performed, however, to identify the sources responsible for the measured variance in TRT.

E. Further explaining the variance in speech reception

It should be noted that the amounts of variance accounted for in the current study may be underestimates. They are attenuated as result of their own nonunity test-retest reliabilities (i.e. their measurement error) and as a result of the measurement error in the independent variables included in the regression model. The correlation coefficient between two variables, r_{XY}, can be corrected for attenuation by applying the formula by Spearman: $r_{XY}' = r_{XY} / \sqrt{r_{XX}\, r_{YY}}$ (see e.g. Allen and Yen, 1979), where r_{XY}' is the de-attenuated correlation coefficient, and r_{XX} and r_{YY} are the test-retest reliabilities of each of the two variables. The amount of explained variance can be corrected for the test-retest reliabilities likewise. Some tests in the current

Proportions	SRT_{STAT}	SRT_{MOD}	MR
Total variance	1.000	1.000	1.000
Explained [Table 4-V]	0.310	0.562	0.472
Explained [de-attenuated]	0.440	0.670	0.664
Unexplained variance	0.560	0.330	0.336

Table 4-VII Overview of the amount of explained and unexplained variance in speech reception in stationary noise (SRT_{STAT}), speech reception in modulated noise (SRT_{MOD}) and masking release (MR).

experiment were only measured once: their test-retest reliabilities were estimated to be 0.90.

Table 4-VII gives an account for the variance in the three measures of speech reception in this study. It can be seen that more than half of the variance in SRT_{STAT} and about one-third of the variance in SRT_{MOD} and masking release is still not accounted for, even after taking the effects of measurement error into account.

One explanation for the large amount of unexplained variance may be that our selection of auditory and nonauditory tests was not extensive enough, even though specifically those tests were selected that were found in literature to be related to speech reception in noise. Another explanation, however, may be found in the fact that the current experiment included measures of i. peripheral auditory processing, and ii. cognitive measures, which were measured visually. Not included were tests that measure cognitive or other supracochlear factors that are specifically auditory in nature. However, as stated in the Introduction, the results of these kinds of auditory tests that measure cognitive abilities are often hard to interpret, because hearing-impairment is always a confounding factor.

Moreover, supra-threshold auditory processing was only measured in the frequency region around 1 kHz. Performing these measurement in a broader frequency range may be expected to improve the amount of explained variance, particularly because some supra-threshold deficits are more prominent in the higher frequencies (George et al., 2006). However, including a broader PTA (adding the threshold at 2 kHz to the average) in the regression model did not improve the current results. Further research will have to resolve whether these suggestions might account for the remaining unexplained variance.

V. FINAL CONCLUSIONS

The main results arrived at in the previous sections can be summarized as follows:

I. Two components could be derived, one auditory and the other nonauditory in nature, that account for the major part of the variance of the auditory and nonauditory measures in the test battery. [Results, Section C]

II. Variance in speech reception is mainly governed by auditory factors. However, nonauditory factors play a significant role when it comes to speech comprehension in nonstationary noise, or to masking release or 'listening in the gaps'. [Results, Section D]

III. In clinical practice, the relevance of auditory and nonauditory factors for speech reception is adequately estimated by the audiogram (PTA), and by the tests of spatial working memory (SWM-B) and perceptual closure in text reception (TRT). [Results, Section E]

IV. The TRT is the only variable that significantly loads on both the auditory and the nonauditory derived components. Moreover, only the TRT is still significantly associated with speech reception (masking release) after age and audiogram have been partialled out. This makes it possible to investigate the sources responsible for deteriorated speech reception in clinical practice. [Results, Section F & Discussion].

V. Even when the effect of attenuation of correlations as a result of measurement error is taken into account, more than half of the variance in SRT_{STAT} and about one-third of the variance in SRT_{MOD} and masking release cannot be accounted for by the current battery of tests. [Discussion, Section E].

VI. ACKNOWLEDGEMENT

This research was supported by the Heinsius-Houbolt Foundation, The Netherlands.

5 The combined effects of reverberation and non-stationary noise on sentence intelligibility

Erwin L.J. George, Joost M. Festen and Tammo Houtgast

Submitted for publication in the Journal of the Acoustical Society of America.

Listening conditions in everyday life typically include a combination of reverberation and non-stationary background noise. It is well-known that sentence intelligibility is adversely affected by these factors. To assess their combined effects, a model is introduced which combines two models of speech perception, the Extended Speech Intelligibility Index (ESII) and the Speech Transmission Index (STI). First, the effects of reverberation on non-stationary noise (i.e. reduction of masker modulations) and on speech quality are evaluated separately. Subsequently, the ESII is applied to predict the speech reception threshold (SRT) in the masker with reduced modulations. To validate this model, SRTs were measured for ten normal-hearing listeners, in various combinations of non-stationary noise and artificially created reverberation. After taking the characteristics of the speech corpus into account, results show that the model accurately predicts SRTs in non-stationary noise and reverberation for normal-hearing listeners. Furthermore, it is shown that, when reverberation is present, the benefit from masker fluctuations may be substantially reduced.

I. INTRODUCTION

Everyday listening situations generally include a combination of reverberation and noise. It is known that these two factors adversely affect speech comprehension. Their effects need to be taken into account when using a model approach to predict speech intelligibility, which can be useful, for example, in the design of public address systems, conference rooms or public facilities in general. Examples of objective speech intelligibility models are the Articulation Index (AI; French and Steinberg, 1947; Kryter, 1962) or more advanced models, based on the AI, like the Speech Intelligibility Index (SII; ANSI, 1997) or the Speech Transmission Index (STI, Houtgast et al., 1980; Steeneken and Houtgast, 1980). These models have been shown to accurately assess the combined effects of reverberation and stationary noise on speech.

However, many everyday backgrounds are non-stationary or 'fluctuating' in nature. Normal-hearing listeners are able to make use of the relatively silent periods or 'gaps' in these non-stationary background to improve their speech intelligibility (Festen and Plomp, 1990; George et al., 2006). The STI and the SII models do not take this effect of background fluctuations into account and, thus, tend to underestimate speech intelligibility. To solve this problem, Rhebergen and Versfeld (2005) introduced the Extended Speech Intelligibility Index (ESII), that is able to accurately predict speech reception thresholds (SRTs) in fluctuating noise for normal-hearing listeners. However, it remained unclear how the adverse effect of reverberation on speech intelligibility, in combination with the presence of non-stationary noise, could be assessed.

Reverberation has a double effect on the intelligibility of speech in non-stationary noise. First, reverberation adversely affects the quality of the speech, an effect that is present even in silence. Secondly, reverberation 'smears out' the temporal waveform of the masker, thus attenuating masker modulations and reducing the size of the 'gaps' in the masker envelope. Therefore, the benefit in speech reception from masker modulations or 'masking release', as often obtained by normal-hearing listeners, will be reduced when reverberation is present.

The current investigation introduces a model that combines the STI and the ESII to predict speech intelligibility in listening conditions with both reverberation and fluctuating noise. To validate the model, SRTs are measured in several non-stationary backgrounds. To further test the assumptions underlying the model, the reverberation is chosen to act either on the speech only, on the noise only, or on both speech and noise, of which the latter condition - of course - most closely resembles situations in daily life. Furthermore, the consequences of reverberation for masking release are investigated.

II. MODEL DESCRIPTION

A. Classic models of speech intelligibility

The Speech Transmission Index (STI; Houtgast et al., 1980; Steeneken and Houtgast, 1980) is based on the observation that speech intelligibility is related to the preservation of the envelope spectrum of speech. Noise and reverberation in the listening environment adversely affect this preservation of modulations. The preservation is characterized by the Modulation Transfer Function (MTF), which quantifies the detrimental effects of distortions on the modulations of the envelope of (band-filtered) speech. The MTF is expressed in matrix-form, by determining the modulation transfer indices m, for each of 14 modulation frequencies (0.63-12.5 Hz) and for each of 7 spectral octave bands (center frequencies 125-8000 Hz). The modulation transfer indices are translated to effective signal-to-noise ratios, averaged over modulation frequencies, and converted to an index between zero and one, after which a weighted average over the octave bands gives the STI.

The STI does not take individual properties of talkers or listeners into account, but is purely a measure of the transmission channel or, to be more specific, the acoustic characteristics of the environment. It has been shown to be strongly related to speech intelligibility (Houtgast and Steeneken, 1984) and is widely used as an acceptability criterion in room acoustics or telecommunication.

The Speech Intelligibility Index (SII; ANSI, 1997) is in fact a very similar model, but its applicability is essentially restricted to assessing the effect of noise. Instead of the modulation transfer functions, the spectra of speech and noise are used to determine the effective signal-to-noise ratio (SNR) in each spectral band. The SNRs are converted to an index between zero and one and taking the weighted average over the spectral bands gives the SII. It is often interpreted as the amount of undistorted speech information available for an individual listener. Recently, the Extended Speech Intelligibility Index (ESII; Rhebergen and Versfeld, 2005) was introduced that makes it possible to apply the SII in non-stationary backgrounds by calculating and averaging the SIIs determined in short time frames.

In contrast with the STI, the SII does take the individual properties of talkers and listeners into account. It is possible to include the individual audiogram of the listener in the calculation, and different spectral weightings can be chosen to assess various talker styles or speech materials.

The SII, like the STI, is related to speech intelligibility: poor intelligibility is associated with SII-values below 0.45 and good intelligibility assumes a SII larger than 0.60. In the assessment of speech intelligibility for individual listeners, the Speech Reception Threshold (SRT) is commonly used, defined as the signal-to-noise ratio that the listener needs to correctly repeat 50% of the presented sentences correctly. For normal-hearing listeners, the SRT corresponds to a STI or SII of about 0.33. This means that, in situations with stationary noise without reverberation, both models predict a SRT of about -5 to -4 dB SNR, consistent with results as found in literature. With the STI-model, the SRT can be predicted for all other combinations of stationary noise and reverberation, always assuming that the STI at the threshold is constant.

In summary, the STI can be used to evaluate the detrimental effects of reverberation and stationary noise on speech intelligibilty, while the recently developed ESII can be applied to predict the effect of non-stationary noise on SRTs. This motivated us to derive a procedure to assess the combined effects of reverberation and non-stationary noise. This approach, based on combining the two models, is outlined in the following section.

B. Combining the STI and the SII

The STI model already combines the effects of noise and reverberation for sentence intelligibility in one model, only with the restriction that the noise should be stationary in character. Let's start with evaluating this model, after which we'll generalize it to non-stationary noise.

Reverberation time T is defined by the time after which a sound has died away to a level 60 dB below its original level. It is commonly determined by extrapolating the slope of the early decay curve. When assuming an exponentially decaying sound field, the STI can be easily calculated for any combination of reverberation time T and speech-to-noise ratio in stationary noise SNR_{STAT} (see Duquesnoy and Plomp, 1980). The modulation transfer index *m,* as used in the STI calculation, is then the product of two terms, one reflecting the effect of reverberation and the other reflecting the effect of background noise. The result is represented in Fig. 5-1 in the form of iso-STI (iso-intelligibility) contours. These contours indicate that it is possible to find, at each value of T, a SNR_{STAT} at which 50% of the presented sentences can be correctly reproduced. Results by Duquesnoy and Plomp (1980), as also displayed in Fig. 5-1, confirm that the SRT in different reverberant sound fields is indeed distributed along an iso-STI contour, i.e. can be represented by one single STI-value. For normal-hearing listeners, the resulting STI is about 0.33, for all combinations of reverberation and stationary noise.

To generalize this approach to non-stationary backgrounds, we need to find the threshold speech-to-noise ratio in fluctuating noise (SRT_{FLUC}) that gives the same intelligibility as obtained at the threshold in stationary noise (SRT_{STAT}). Thus, the SII or STI at the threshold should be the same, whether the noise is stationary or fluctuating in character. Rhebergen and Versfeld (2005) showed that the extended SII may be applied to evaluate the SII in fluctuating noise. Thus, we apply the ESII to determine the speech-to-noise ratio in fluctuating noise (SRT_{FLUC}), that, in terms of

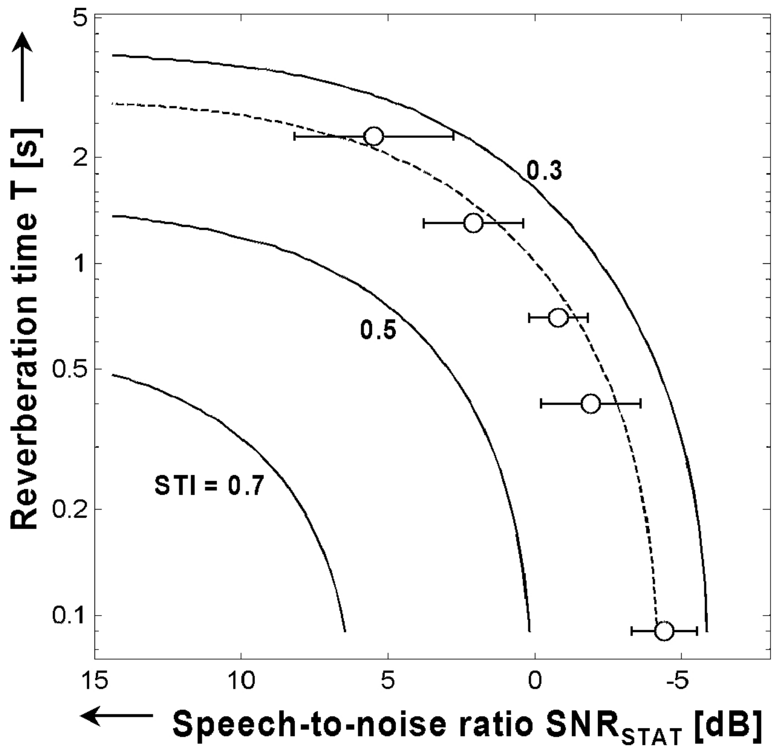

Figure 5-1 Iso-STI contours for listening conditions that include a combination of reverberation and stationary noise. The data points are measurement results for normal-hearing listeners by Duquesnoy and Plomp (1980), representing the signal-to-noise ratio at the SRT, i.e. the level at which 50% of the sentences could be correctly reproduced, for various reverberation times T. Redrawn, from Houtgast et al., 1980.

intelligibility, is equivalent to the signal-to-noise ratio in stationary noise (SRT_{STAT}).

However, this approach to predict SRTs in fluctuating noise assumes that reverberation and noise independently affect speech reception. This assumption is correct in the case of stationary noise: the reverberation does not affect the characteristics of the background noise, since reverbed stationary noise is still

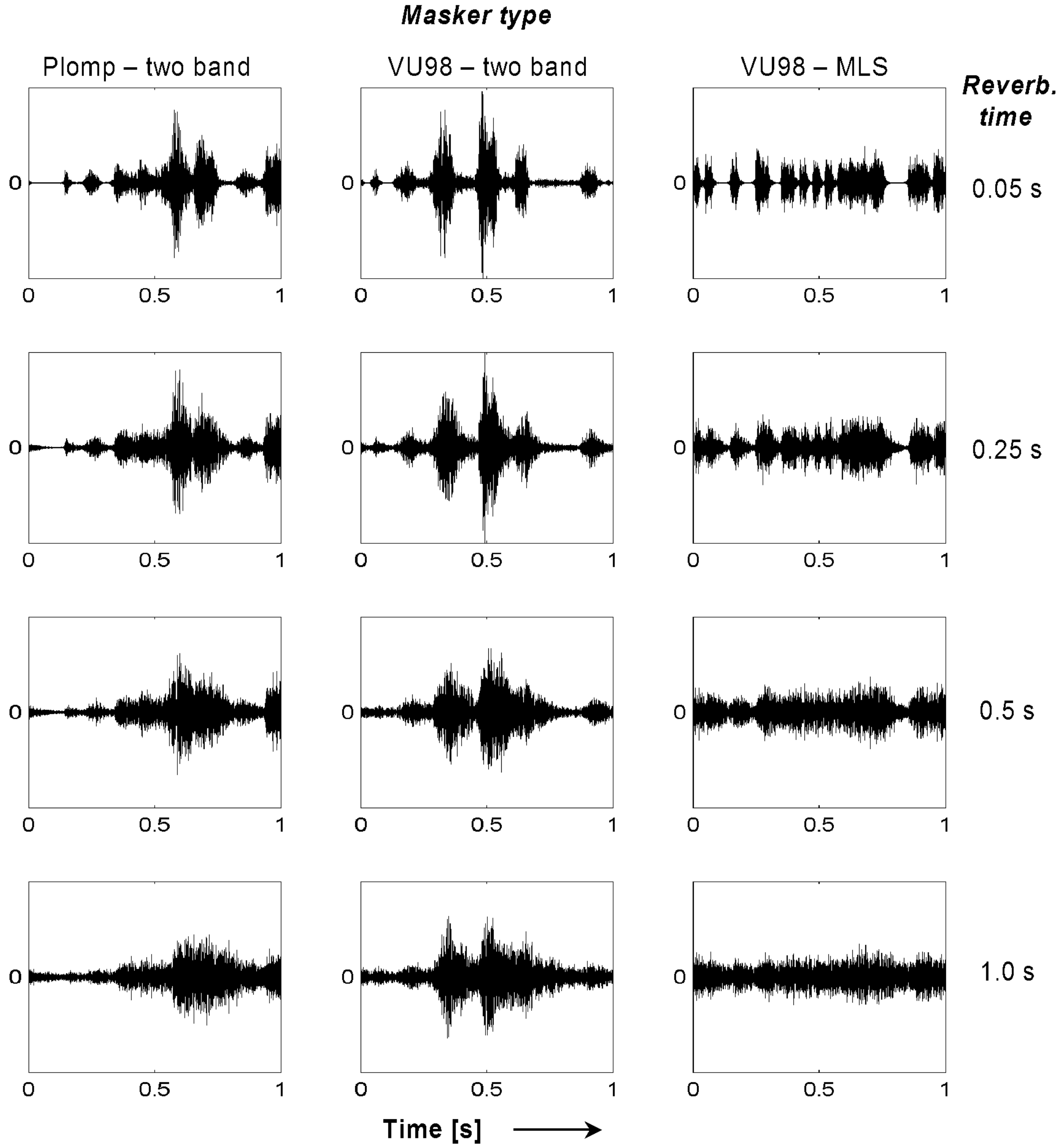

Figure 5-2 The effect of simulated reverberation on the temporal waveforms of three different types of non-stationary maskers. To be able to compare the masker types in terms of visible waveform modulations, only the modulations for one representative octave band (around 2 kHz) are shown. Details on the masker types and the reverberation procedure can be found in the Methods Section.

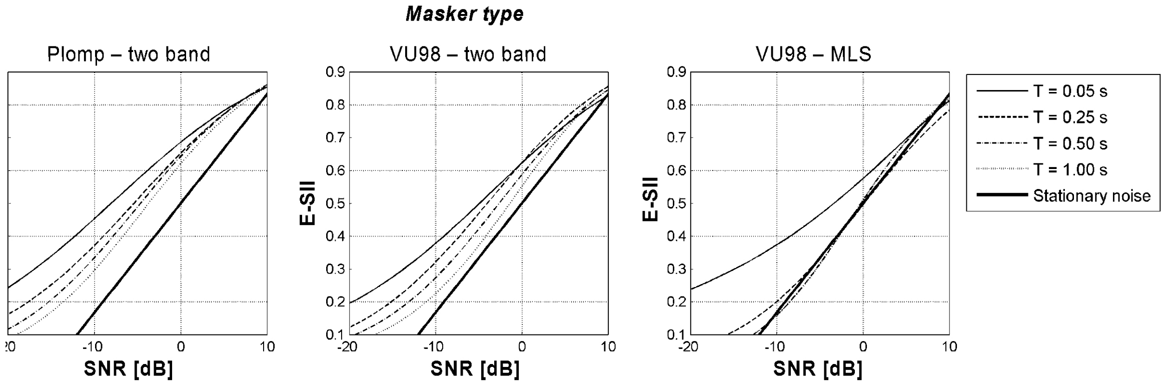

Figure 5-3 Extended Speech Intelligibility Index (ESII) as a function of speech-to-noise ratio (SNR) for three different types of non-stationary maskers with reverberation time as a parameter. ESII-values were calculated for each reverbed masker and fitted to the three-parameter asymmetric logistic function $ESII(\mu,\sigma,n)=[1+n^*\exp(\mu\text{-}SNR)/\sigma]^{-1/n}$, which reduces to a symmetrical sigmoid for n=1. For stationary noise, ESII is simply the linear function (SNR+15)/30, cf. ANSI, 1997. Details on the masker types and the reverberation procedure can be found in the Methods Section.

stationary in nature. When non-stationary noise is concerned, however, the reverberation will not only adversely affect the speech, but will also change the characteristics of the noise. Some examples of the effect of reverberation on the temporal waveform of the noise are displayed in Fig. 5-2, for three different types of maskers.

Reverberation was simulated here by convolving the non-stationary noise with synthetic impulse responses, the details of which are described in the Methods Section. As shown, the relatively silent periods or ‘gaps’ in the maskers are reduced in size by the reverberation. At larger reverberation times, masker fluctuations are further reduced and eventually, there are no fluctuations or ‘gaps’ present anymore and the fluctuating noise has become more-or-less stationary in character. A direct consequence of this effect is that the benefit from masker fluctuations in speech reception, as often observed for normal-hearing listeners, is reduced as reverberation time increases, until the point where no masking release is obtained anymore. Put differently, reduction in masker fluctuations as a consequence of reverberation directly affects the SRT in fluctuating noise.

The effect of reverberation on non-stationary noise for speech intelligibility can be quantified by applying the ESII, as shown in Fig. 5-3. This figure displays, for various noise types and reverberation times, the calculated amount of available speech information, expressed by the ESII, as a function of signal-to-noise ratio. As

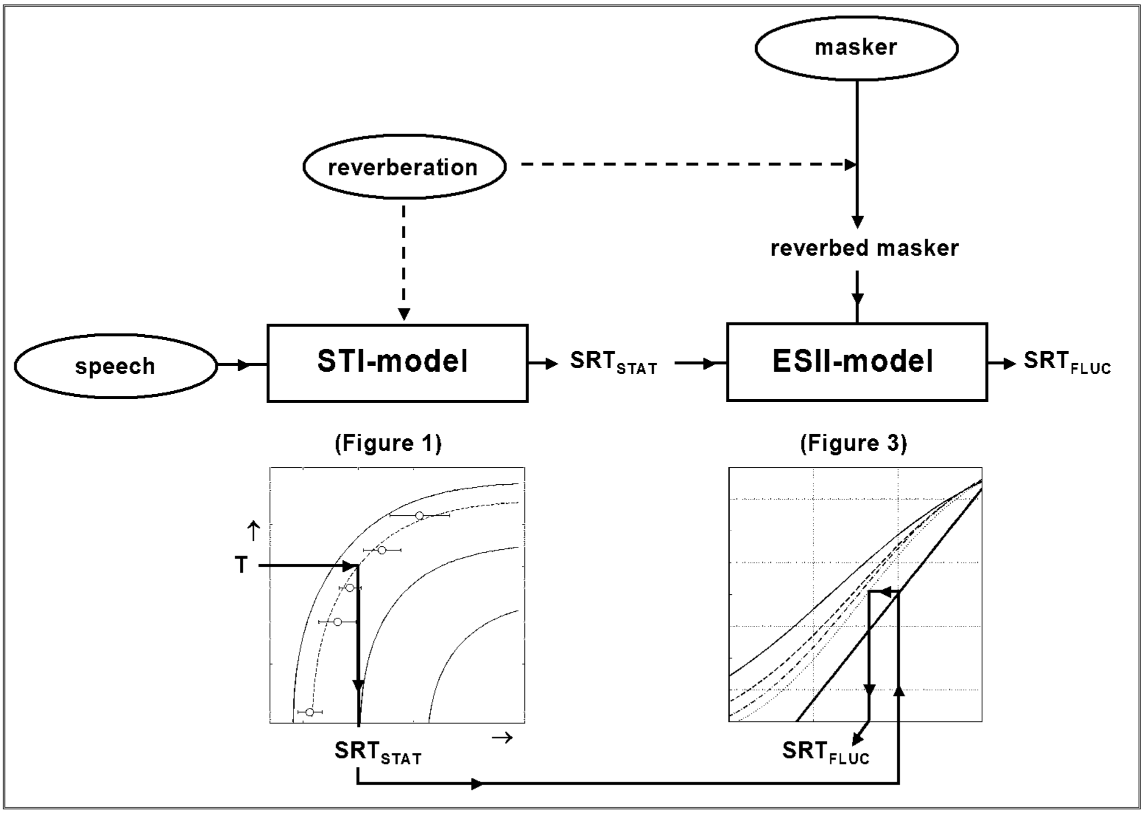

Figure 5-4 Overview of the model. The effect of reverberation on speech is described by the STI-model, while the effect of reverberation on the masker is also evaluated separately. Subsequently, the ESII is applied to determine the combined effects of both reverberation and masking noise on sentence intelligibility. The shown example applies to a reverberation time of T = 1 second and the 'Plomp - two band' masker type.

reverberation time increases, the curves relating intelligibility (the ESII) to signal-to-noise ratio for fluctuating maskers approach the curve for stationary noise. This effect is most prominent for the masker in the right-most panel, for which the T = 0.5 s curve can hardly be distinguished from the curve for stationary noise anymore. Choosing the appropriate reverberation time and noise type, the resulting curve can be used to translate the speech-to-noise ratio in stationary noise (SRT_{STAT}) to an SNR in non-stationary noise (SRT_{FLUC}) that is equivalent in terms of the effect on intelligibility.

In summary, Fig. 5-4 gives an overview of the model that is used for predicting SRTs in combinations of non-stationary noise and reverberation. The STI-model, in the form of iso-STI-curves, is applied to determine the effect of the reverberation T on speech quality (see Fig. 5-1). This gives rise to a prediction of the SRT in stationary noise, SRT_{STAT}. The effect of reverberation on the masker is also separately evaluated. Finally, the ESII is used (see Fig. 5-3) to translate the obtained SRT_{STAT} to an equivalent SRT_{FLUC} in the reverbed fluctuating noise.

To determine whether the proposed model, combining the STI and the ESII, gives accurate predictions of speech intelligibility, an experiment was conducted, described in the next section, in which SRT-measurements for various combinations of noise types and reverberation times were performed.

III. EXPERIMENT AND METHOD

A. Participants

Ten young, normal-hearing listeners participated in this experiment. Eight were students from the VU University, while the other two were young university graduates. They reported no problems with their hearing or with speech reception, and were selected to have pure-tone hearing thresholds equal to or better than 10 dB HL at octave frequencies between 0.25 and 4.0 kHz. Their ages ranged from 18.3 to 31.4 years, with an average of 22.9 years.

B. Method

SRT-measurements were conducted by using a simple adaptive up-down procedure as described by Plomp and Mimpen (1979). In each condition, the masker and a list of thirteen sentences, unknown to the listener, were presented. At a constant reverberation time, the speech-to-noise ratio was varied adaptively to estimate the SRT. Lower SRT-values indicate better performance. In each condition, the first sentence was presented at a level below threshold and repeated, at 4-dB higher levels with each repetition, until the listener was able to reproduce it correctly. The remaining twelve sentences in the list were presented only once, following an one-up-one-down adaptive procedure, with a 2-dB step size. An errorless reproduction of the entire sentence was required for a correct response. The SRT was estimated as the average presentation level of sentences 4 to 13.

C. Stimuli

1. Sentences

Two sets of short meaningful sentences were used as speech material. The first speech corpus was developed and evaluated by Plomp and Mimpen (1979), consisting of ten lists of thirteen Dutch sentences, uttered by a female speaker. All ten lists were used in our experiment. The second corpus used here is the ‘VU98’ corpus (Versfeld et al., 2000), which consists of 39 lists of sentences of a male talker and an equal number of lists for a female talker. The first twenty lists by the female speaker were used. Both corpora were developed to enable efficient measurement of SRTs in noise and are considered as mutually equivalent, although the speaking style of the VU98 sentences is generally more informal.

2. Maskers

Sentences were presented in three different maskers; all maskers were designed to mimic the intensity fluctuations of speech. However, their specific characteristics were different. The first masker was generated using the method described by Festen and Plomp (1990), which splits up a stationary masker in a low- and high-frequency part with 1000-Hz crossover frequency. Both parts are then modulated separately with the envelope of speech from the corresponding frequency region, after which they are added while restoring the original level-ratio between the two bands. The second masker was similar to the first; the only difference was that the first masker was based on modulations in the speech corpus of Plomp and Mimpen (1979), while the second masker was based on the speech corpus of Versfeld et al. (2000). The third masker was constructed by multiplying a stationary noise with a time-scaled maximum-length sequence (MLS) with a value of either zero or unity. After scaling, this resulted in an on-off noise with a flat modulation spectrum between approximately 3 and 20 Hz, thus covering the spectrum of all the important envelope modulations in speech (Houtgast and Steeneken, 1985; Drullman et al., 1994). The first two maskers are similar in nature and will be denoted as the 'two band' masker type, while the third masker is denoted as the 'MLS' type. For each of the three maskers, ten-second samples were constructed, from which a sequence of about two to three seconds was randomly chosen for each sentence presentation. Examples of the waveforms of all three maskers are displayed in Fig. 5-2.

Each of the three maskers was combined with the corresponding speech corpus: the Plomp & Mimpen speech corpus was presented in masker one, while the VU-lists were used to measure SRTs in the second and the third masker. This approach made it possible to assess the influence of speech corpus and masker type on model predictions. Since the characteristics of the first two maskers were essentially the same, the difference in model predictions between the two expresses the effect of speech corpus on model predictions. The influence of masker type may be assessed by comparing SRT-predictions for the second and the third masker.

3. Reverberation

Reverberation was introduced by convolving the auditory signals with synthetic impulse responses. This approach made it possible to define the reverberation time in a systematic way, excluding unwanted effects of room acoustics which may be present when using real, recorded impulse responses. In terms of the modulation transfer function, the artificial reverberation was identical to purely exponentially decaying real reverberation of the same reverberation time. The impulse responses were created by subjecting white noise to pure exponential decay. The slope of the decay and the length of the impulse response were determined by the desired reverberation time T, chosen to be 0.05, 0.25, 0.5 and 1.0 seconds. These impulse responses of various reverberation times were multiplied with either only the speech, or only the masker, or both the masker and the speech. This leads to three measurement conditions for each reverberation time. The short reverberation time of 0.05 was included as a reference condition and expected to be similar to ‘no reverberation’. At this shortest T, reverberation was chosen to always act on both the speech and the noise. This gives a total of ten ‘reverberant’ conditions: three modes of reverberation times three reverberation times, plus the reference condition.

In all conditions, the long-term spectra of the speech and the masker were similar in shape. Reverberation was applied to the speech and / or to the selected masker, after which they were mixed according to the desired signal-to-noise ratio, based on their rms-levels. The spectrum and presentation level of the resulting signal was adapted to reach octave masker levels equal to the middle of the dynamic range for each listener. The lower limit of the dynamic range was chosen to be the individual pure-tone threshold, while the upper limit was the uncomfortable loudness level (UCL), here chosen at 110 dB SPL for all listeners. This approach is commonly used in our laboratory to assure optimal audibility for all listeners.

D. General method and instrumentation

A test session always started with the measurement of the listener’s audiogram. Subsequently, a total of thirty SRT-measurements were performed for each listener, in three blocks of ten SRTs, using one specific masker type within each block of ten ‘reverberant’ conditions. Characteristics of the three maskers and conditions were described above. The order in which the three blocks were presented was randomly determined for each participant. Within each block, confounding of measurement order and sentence lists with condition effects was avoided by counterbalancing the order of conditions across subjects, according to a ten-by-ten digram-balanced Latin square, while list order was kept fixed.

The experiment was run on a Dell personal computer, equiped with a Creative Labs Audigy external sound device and Beyer Dynamic DT48 headphones. Sound calibrations were performed with a Brüel & Kjær Artificial Ear (type 4152) and a Brüel & Kjær 2260 Observer conform ISO 389 (1991). All measurements were performed while the listener and the investigator were seated in a sound-insulated room. SRTs were conducted monaurally, using the participant's best ear, which was chosen according to his or her audiogram, or, in case of equal audiograms, personal preference in telephone conversation.

IV. RESULTS

Measurement results are displayed in Table 5-I for each of the three maskers used in the current experiment. Model predictions were derived by assuming a constant STI of 0.33 at the threshold, i.e. the 'STI at the SRT' = 0.33 for all conditions. The model

		Reverberation time T [s]							
		0.05		0.25		0.5		1.0	
		M	*S*	M	*S*	M	*S*	M	*S*
Plomp - two band	Both reverbed	-15.4	*3.1*	-9.0	*1.5*	-5.3	*2.5*	-0.6	*2.4*
	Only noise	--	--	-13.1	*2.0*	-11.2	*2.2*	-8.6	*1.9*
	Only speech	--	--	-13.8	*2.2*	-11.8	*3.0*	-8.7	*2.1*
VU98 - two band	Both reverbed	-7.8	*2.5*	-2.5	*2.4*	+1.6	*1.5*	+5.1	*2.2*
	Only noise	--	--	-6.9	*1.8*	-6.0	*1.6*	-4.6	*1.7*
	Only speech	--	--	-5.8	*2.1*	-3.2	*2.0*	+3.8	*4.0*
VU98 - MLS	Both reverbed	-13.5	*4.2*	-1.1	*2.0*	+2.1	*1.7*	+6.0	*1.9*
	Only noise	--	--	-3.9	*1.5*	-3.5	*0.7*	-3.4	*0.9*
	Only speech	--	--	-13.1	*3.9*	-8.3	*4.4*	-2.1	*7.1*

Table 5-I Group means (M) and standard deviations (*S*) for speech reception thresholds (SRT, in dB SNR) in ten reverberant conditions, for three combinations of speech corpus (Plomp and VU98) and non-stationary masker type ('two band' and 'MLS').

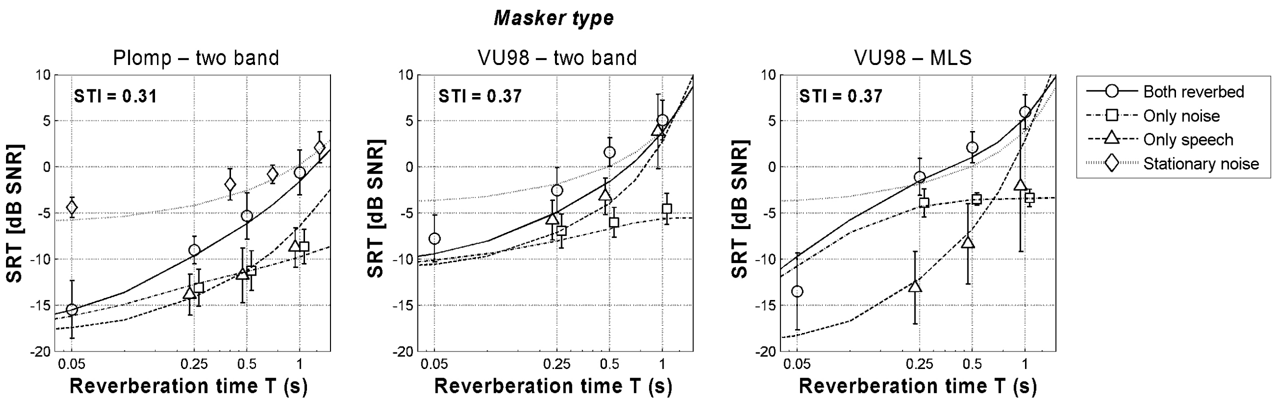

Figure 5-5 Speech reception thresholds (SRTs) as a function of reverberation, for three combinations of speech corpus (Plomp and VU98) and non-stationary masker type ('two band' and 'MLS'). Model predictions, represented by curves, were determined by applying a STI-model with 18 modulations bands, as suggested by Van Wijngaarden and Houtgast (2004). For each speech corpus (Plomp or VU98), the STI at threshold was chosen to give an optimal fit ('least squares') between measurements and model predictions. The leftmost panel also displays results from SRT-measurements in stationary noise, taken from Duquesnoy and Plomp, 1980.

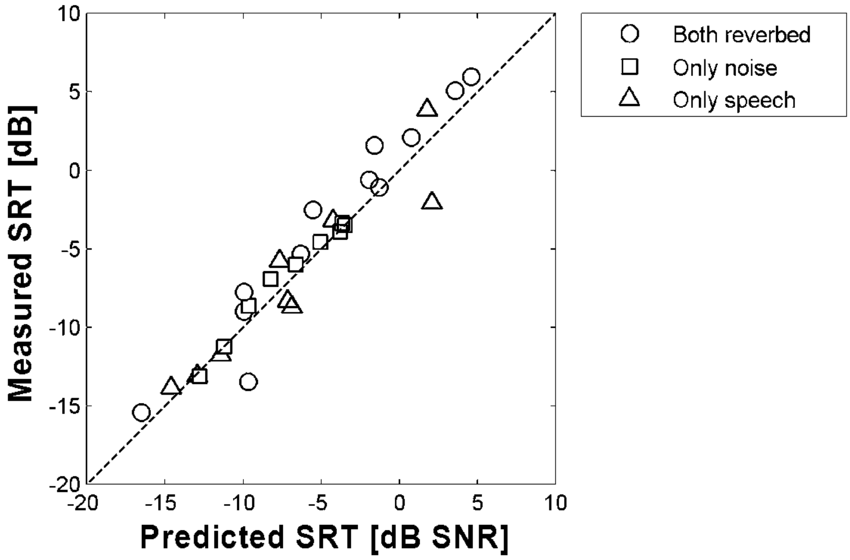

Figure 5-6 Scatter plot of the observed speech reception thresholds (SRTs) and the predicted SRTs, for all combinations of speech corpus, masker type and reverberation.

predictions appeared reasonable for the Plomp & Mimpen speech corpus, but the SRTs measured with the VU corpus were not adequately described. In all cases where the VU corpus was used, listeners obtained higher SRTs than predicted, i.e. needed more undistorted speech information to correctly reproduce half of the sentences.

These findings can be understood when considering results by Van Wijngaarden and Houtgast (2004). They showed that the classic STI, as applied to calculate the model predictions, indeed underestimates the adverse effect of reverberation on speech intelligibility when conversational speech by an untrained talker is concerned. This effect was explained by the relatively stronger contributions of higher modulation frequencies in this type of speech material. They obtained better results for an adapted version of the STI-model, in which the modulation frequency range was extended to 31.5 Hz, including a total of eighteen modulation frequency bands instead of the classic fourteen. The difference in speaker styles is then expressed by differences in the 'STI at the SRT' between the speech corpora: 0.31 for the Plomp & Mimpen corpus and 0.37 for the VU corpus.

Figure 5-5 shows our measurement results and model predictions when the differences in speaker style are taken into account, that is, by applying the eighteen band STI model to determine the effect of reverberation on speech. Model predictions, represented by curves, were derived here by choosing the STI at the threshold, for each speech corpus, such that an optimal fit was obtained with the measurements.

Figure 5-6 displays a scatter plot of the observed SRTs and the model predictions, for all the conditions from the current experiment. It can be seen that the observed SRTs and the model predictions are in good agreement. The standard deviation of the data from the fitted model is 1.7 dB, while the maximum deviation is 4.2 dB.

V. DISCUSSION

A. Improving model predictions

An optimal fit was obtained between observed and predicted SRTs by choosing a STI at the threshold, for each speech corpus (Plomp or VU98). This fitting procedure gave rise to a STI of 0.31 for the Plomp-speaker and a STI of 0.37 for the VU98-speaker. These values are consistent with the results by Van Wijngaarden and Houtgast (2004) and the difference between them can be accounted for by the effect of speaker style.

For the VU98 speech corpus, the chosen STI, giving rise to an optimal fit between data and model predictions, was based on data from both the VU98 / 'two band' masker combination and the VU98 / 'MLS' masker combination. The fit could be further improved by distinguishing between the two combinations, i.e. by fitting an optimal STI-value for each combination separately. In that case, a STI-value of 0.35 was obtained for the VU98 / 'MLS' masker combination, while for the combination of the VU98-speaker and the 'two band' masker, an optimal fit was obtained using a

STI-value of 0.40. The standard deviation of the data improved from 1.7 dB to 1.3 dB, while the maximum deviation decreased to 2.6 dB.

The obtained STI-value of 0.35 for the VU98 / ‘MLS’ masker combination is still consistent with the results of Van Wijngaarden and Houtgast. In contrast, the obtained STI-value of 0.40 for the combination with the ‘two band’ masker seems rather high. Apparently, listeners find it hard to understand speech in this specific combination of masker type and speaker style, giving rise to an elevated STI-value. However, the current data set is too small to be able to further investigate the processes underlying this finding.

B. Effect of reverberation on speech quality and on non-stationary maskers

It can be seen in Fig. 5-5 that, in all conditions, more reverberation gives rise to a higher SRT. This was to be expected, since, as mentioned earlier, reverberation reduces the quality of the speech, and also reduces the modulations in the masker. The figure also shows that, as expected, the curves for the conditions in which only the noise was reverbed approach the limit of the SRT in stationary noise (around -5 to -4 dB SNR), for all three masker types. The curves for the conditions in which only the quality of the speech was affected by reverberation theoretically reach an SRT of +15 dB at reverberation times around 2 to 3 seconds. In that case, reverberation affects the quality of speech so severely that the masker can no longer be tolerated.

The curves for the conditions where both the noise and the speech were reverbed can be regarded as ‘weighted sums’ of the other two curves. In the left panel and especially in the middle panel, the SRT in these conditons is mainly affected by the adverse effect of the reverberation on the quality of speech, while the effect of reduced masker modulations on intelligibility remains reasonably limited. For the rightmost panel - the ‘MLS’ masker - the situation is the other way around: the effect of reduced speech quality is relatively small for most reverberation times (up to 0.8 seconds) and the effect of the reverberation on the masker is dominant in affecting the SRTs. Even when only a small amount of reverberation is introduced ($T = 0.05$ s), the modelled effect of the reverbed noise on speech intelligibility is substantial. Data from a pilot experiment in our department (unpublished) showed an average SRT of -20.8 dB SNR (standard deviation 2.7 dB) in this type of masker when no reverberation is present, while the currently measured SRT at $T = 0.05$ s is -13.5 dB SNR, a difference of about 7 dB.

The fact that even a small amount of reverberation has such a large effect on SRT in the 'MLS' masker can be explained by the masker's modulation spectrum. Since its modulation spectrum is flat, the 'MLS' masker contains a relatively large amount of fast modulations, up to 20 Hz. In addition, the modulation depth of the 'MLS' masker was 100%. These fast and deep masker fluctuations are more vulnerable to reverberation (more easily 'smeared') than modulations with lower frequencies and smaller modulation depths, as mostly present in the 'two band' masker type.

C. Effect of reverberation on masking release

Finally, consider the difference between the predictive curves for the conditions where both the noise and the speech were reverbed and the curves for stationary noise (Fig. 5-5). This difference gives an estimate for the obtained masking release, that is, the benefit in speech reception due to masker modulations, at a specific reverberation time. At reverberation times below 0.1 seconds, masking release is clearly present (up to 10 dB), cf. Festen and Plomp (1990) or George et al. (2006). When reverberation increases, masking release is still present up to reverberation times near 1.0 second in the first panel. However, in the second panel, masking release is reduced substantially, becoming almost zero at a reverberation time around 0.5 seconds. In the rightmost panel, a reverberation time of 0.25 seconds is already enough to almost fully eliminate the obtained benefit. Thus, when the effect of reverberation on the character of the noise and the quality of the speech is taken into account, the benefit that listeners obtain from masker modulations is largely reduced.

These differences between the panels concerning the reduction of masking release can be understood when considering the effects of reverberation on the speech quality or on the masker, as explained in the previous Section. However, the reduction of masking release by reverberation in non-stationary noise may have large consequences. In everyday life, the style of the speaker is often not very clear and fairly informal, i.e. comparable or even worse in quality than the speaker used in the middle and right panels of Fig. 5-5. Moreover, reverberation times around 0.4 to 0.5 seconds, or even larger, are very commonly found in daily listening environments like living rooms, offices, or hospitals. The results from the current experiment indicate that masking release in even these common everyday situations may be substantially reduced or even absent.

It should be noted, however, that the presented model predictions only hold when the listener is in the indirect sound field, that is, for distances to the speaker larger than about 0.2 $(V/T)^{1/2}$. For a typical living room of 75 m^3 and T = 0.5 s, this distance is 2.5 meters (see Duquesnoy and Plomp, 1980). When the listener approaches the speaker more closely, the effect of reverberation is reduced and the obtained masking release in fluctuating noise is likely to be larger than predicted by the model. Nevertheless, when the listener is not very close to the speaker, the obtained benefit from masker modulations may be considerably reduced in situations with noise and reverberation.

VI. CONCLUSION

A model has been introduced that combines the Extended Speech Intelligibility Index (ESII) and the Speech Transmission Index (STI) to assess the combined effects of reverberation and non-stationary noise on speech intelligibility. After taking the characteristics of the speech corpus into account, the model accurately predicts the SRT in non-stationary noise and reverberation for normal-hearing listeners. Further consideration of the model predictions shows that the masking release, as observed in the absence of reverberation, may be substantially reduced in everyday listening situations with noise and reverberation.

VII. ACKNOWLEDGEMENT

This research was supported by the Heinsius-Houbolt Foundation, The Netherlands. Thanks are due to Johannes Lyzenga for his contribution to the discussions underlying this paper and for providing the 'MLS' masker.

Measuring the effects of reverberation and noise on sentence intelligibility for hearing-impaired listeners

6

Erwin L.J. George, Joost M. Festen and Tammo Houtgast

To be submitted for publication in the Journal of the Acoustical Society of America.

The Speech Transmission Index (STI) is commonly used to quantify the adverse effects of reverberation and stationary noise on speech intelligibility for normal-hearing listeners. Duquesnoy and Plomp (1980; J. Acoust. Soc. Am. 68, 537-544) showed that the STI can also be applied for presbycusic listeners, relating the Speech Reception Thresholds (SRTs) in various reverberant conditions to a fixed, subject-dependent STI-value. The current paper extends their results to a wider range of hearing-impaired listeners. Moreover, a reverberant analogue of the SRT is presented, which determines the amount of reverberation that a listener can sustain to still understand 50% of the presented sentences. Results show that this Speech Reception Reverberation Threshold (SRRT) can be consistently determined for both normal-hearing and hearing-impaired listeners. Moreover, the STI indeed appears to be a convenient, single number to quantify speech recognition performance of hearing-impaired listeners in situations with noise and reverberation. The nature of the hearing loss (presbycusic, congenital, or otherwise) does not affect this result. Finally, it is suggested that determining the effect of reverberation on sentence intelligibility directly, using the SRRT, may be helpful in further understanding the supposed importance of cognitive processing in reverberant listening conditions.

I. INTRODUCTION

Understanding speech in everyday life can be a real challenge: the common presence of background noise and reverberation in everyday listening situations seriously affects the reception of conversational speech. Normal-hearing listeners are usually able to tolerate moderate amounts of noise or reverberation, but hearing-impaired listeners often complain of not being able to understand what is being said, even when the level seems sufficient.

Plomp and Mimpen (1979) developed a standardized method to determine the Speech Reception Threshold (SRT), which measures the signal-to-noise level that a listener can sustain while still being able to correctly reproduce 50% of the presented sentences. The SRT has been widely applied for normal-hearing and hearing-impaired listeners to assess the effect of noise on sentence intelligibility (see e.g. Festen and Plomp, 1990; George et al., 2006).

The effect of reverberation on sentence intelligibility can be quantified by applying the Speech Transmission Index or STI (Houtgast et al., 1980; Steeneken and Houtgast, 1980). The STI is based on the observation that both reverberation and noise reduce the envelope fluctuations in speech. Through the modulation transfer function, the influence of reverberation can be translated into an equivalent speech-to-noise ratio, after which it is treated in essentially the same way as additive noise. For normal-hearing listeners, the resulting STI is highly related to sentence intelligibility. Thus, a reduction in STI always gives rise to a corresponding reduction in sentence intelligibility (Houtgast and Steeneken, 1984), irrespective of whether the speech degrading factor is reverberation, noise or spectral filtering.

Measurements by Duquesnoy and Plomp (1980) confirm this result for a group of listeners suffering from presbycusis. They measured SRTs at various reverberation times and showed that, for each combination of noise and reverberation, the 'STI at the SRT' was always the same listener-specific value. This means that individual performance in conditions with noise and reverberation can be predicted when the threshold in noise is known. For normal-hearing listeners, the resulting STI was about one third. Hearing-impaired, presbycusic listeners needed larger STI-values to reach 50% sentence intelligibility, which is a direct consequence of their hearing loss for speech-in-noise (Plomp, 1978).

The main motivation for the current paper was to reproduce and extend these results, specifically concerning two yet unaddressed issues. First, the STI-method of assessing the effect of reverberation (i.e. translating it into an equivalent signal-to-noise ratio) does not necessarily apply to other groups of hearing-impaired listeners. Individual listeners may vary widely in their susceptibility to noise and reverberation (Nabelek and Mason, 1981) and the mechanisms underlying the reception of speech in noise or in reverberation may even be different (Nabelek and Dagenais, 1986; CHABA, 1988).

Secondly, assessing the effect of reverberation on speech intelligibility by measuring the SRT in noise at various reverberation times (like Duquesnoy and Plomp, 1980) always gives rise to a confounding effect of the presence of noise. To avoid this confounding effect, one could measure SRTs or word-intelligibility scores at various reverberation times without background noise (Helfer and Wilber, 1990; Sato et al., 2007). However, it should be possible to develop a test that enables a more efficient examination of solely the effect of reverberation on speech intelligibility.

The current paper extends the results of Duquesnoy and Plomp (1980) by assessing the effect of reverberation and noise on speech intelligibility in a wider range of hearing-impaired listeners. Is the STI, as known from a measured SRT in noise only, indeed a 'convenient single number to systematically quantify the combined effects of noise and reverberation in a hearing-impaired listeners with diverse etiology ? Moreover, a simple adaptive test is presented, similar to the SRT, that was designed to determine the amount of reverberation that an individual listener can sustain and still understand 50% of the presented sentences. In experiment 1, the results of this Speech Reception Reverberation Threshold (SRRT) test are evaluated for a relatively small group of normal-hearing participants. In experiment 2, the test is applied to a larger, diverse group of hearing-impaired participants.

II. EXPERIMENT 1: NORMAL-HEARING LISTENERS

A. Method

1. The Speech Transmission Index

Thresholds for sentence intelligibility were measured in five different conditions, in which sentences were presented in either background noise, or reverberation, or both. Moreover, the conditions differ in whether the signal-to-noise ratio, or the applied reverberation, or both are varied adaptively. The five conditions can be mutually compared by applying the Speech Transmission Index (STI; Houtgast et al., 1980). As a clarification, an overview of the measurement conditions is given in Fig. 6-1.

When assuming an exponentially decaying sound field, the STI can be easily calculated for any combination of reverberation time and speech-to-noise ratio in stationary noise (see Duquesnoy and Plomp, 1980). The result is represented in Fig. 6-1 in the form of iso-STI (i.e. iso-intelligibility) contours. Results by Duquesnoy and Plomp (1980), as also displayed in Fig. 6-1, confirm that SRTs in different reverberant sound fields are indeed distributed along an iso-STI contour, i.e. can be represented by one single STI-value.

In the SRT test, no reverberation is present and the speech-to-noise ratio is adaptively varied, represented by the double arrow along the horizontal axis in Fig. 6-1 (A). Similarly, when there is no noise and the reverberation time is adaptively varied (in the SRRT, see below), the measurement can be regarded as taking place along the logarithmic vertical axis (B). Three other measurement conditions (C, D and E) are shown, which will all be discussed below.

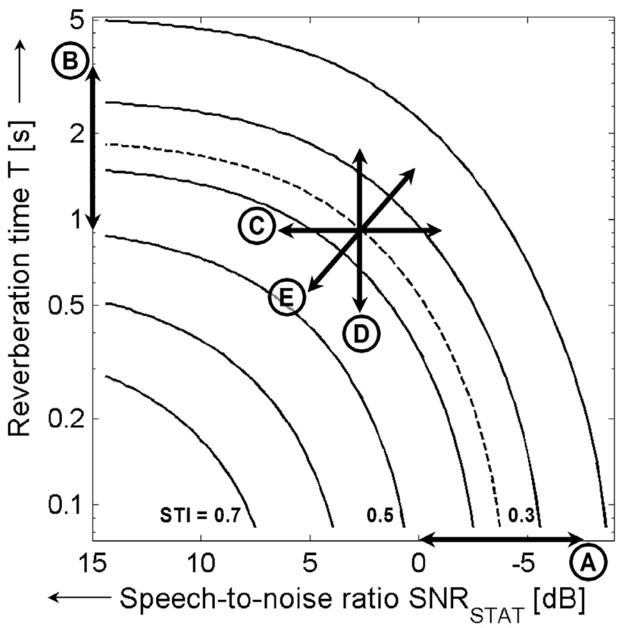

Figure 6-1 Iso-STI contours for listening situations that include combinations of reverberation and stationary noise, plus an overview of measurement conditions: A) Speech Reception Threshold (SRT) without reverberation; B) Speech Reception Reverberation Threshold (SRRT) without background noise; C) SRT with constant reverberation time, T = 0.9 seconds; D) SRRT with constant speech-to-noise ratio, SNR = +2.8 dB; E) Diagonal condition, in which T and SNR are both adaptively varied. The dashed line represents listening conditions for the expected speech intelligibility thresholds, where STI = 0.36.

The STI model applied here is a modification to the original version by Houtgast et al. (1980), as suggested by Van Wijngaarden and Houtgast (2004). They showed that the classic STI underestimates the adverse effect of reverberation on speech intelligibility when informal, conversational speech is concerned. This motivated them to introduce a modified version of the STI-model, including a total of eighteen modulation frequency bands instead of the classic fourteen. Different speaker styles then give rise to a different STI-value at the 50% sentence intelligibility point. For the Plomp & Mimpen corpus, as used by Duquesnoy and Plomp (1980), the resulting STI is about 0.30, for normal-hearing listeners. For the sentences used in the current experiment (Versfeld et al., 2000), a 'STI at the threshold' around 0.36 is expected, represented in Fig. 6-1 by the dashed contour.

2. Noise only (SRT)

The intelligibility of sentences disrupted by noise only was measured with the Speech Reception Threshold (SRT) test, following a simple adaptive up-down procedure as described by Plomp and Mimpen (1979). In each condition, the masker and a list of thirteen sentences, unknown to the listener, were presented. The speech-to-noise ratio was varied adaptively to estimate the SRT, defined as the speech-to-noise-ratio at which 50% of the sentences could be reproduced without error. In each condition, the first sentence was presented at a level below threshold and repeated, at 4-dB higher levels with each repetition, until the listener was able to reproduce it correctly. The remaining twelve sentences in the list were presented only once, following a one-up-one-down adaptive procedure, with a 2-dB step size. An errorless reproduction of the entire sentence was required for a correct response. The SRT was estimated as the average presentation level of sentences 4 to 13.

3. Reverberation only (SRRT)

To determine the intelligibility of sentences disrupted by reverberation only, the Speech Reception Reverberation Threshold (SRRT) test was applied, following an adaptive procedure similar to the SRT. In this condition, there was no masker present during the test, only reverberant speech. The amount of applied reverberation was varied adaptively, to estimate the SRRT, defined as the reverberation time at which 50% of the sentences could be reproduced without error.

The step size of the SRRT was chosen to correspond to the step sizes of the standard SRT test (i.e. 4 dB for the first sentence and 2 dB for all other sentences). In terms of the STI, changing the signal-to-noise ratio with 4 or 2 dB is equivalent to changing the reverberation time with a factor of about 2 or √2, as was derived from Fig. 6-1. More precisely, a factor √2 gives rise to a change in STI equivalent to 1.8 to 2.0 dB per step, for reverberation times between 0.25 and 3 seconds. The first sentence in the SRRT test was presented repeatedly, with reverberation time decreasing with a factor 2 with each repetition, while the remaining twelve sentences were presented only once, adaptively changing the reverberation time with a factor √2 (one-up-one-down) after each sentence. The SRRT was estimated as the geometric average reverberation time while presenting sentences 4 to 13.

4. Noise and reverberation

There were three measurement conditions in which both noise and reverberation were present, each of which is displayed in Fig. 6-1. In condition C, which will be denoted as ‘horizontal’, the speech-to-noise ratio was adaptively varied, while keeping reverberation time constant. In condition D, which will be denoted as ‘vertical’, it was the other way around: the speech-to-noise ratio was kept constant, while adaptively varying the applied reverberation time. The choices for the values of the reverberation time in condition D and the signal-to-noise ratio in condition C were based on the expected ‘STI as the threshold’, as can be derived from Fig. 6-1.

In condition E, both the signal-to-noise ratio and the applied reverberation time were varied, in such a way that the adaptive procedure followed a track perpendicular to the STI-contours. Following this ‘diagonal’ track, under a 45 degree angle with the horizontal and the vertical axis, ensures that the adverse effects of noise and reverberation on speech intelligibility are equivalent in terms of STI. Thus, this measurement condition is the exact compromise between assessing the effect of noise (condition A), on the one hand, and assessing the effect of reverberation (condition B) on speech intelligibility, on the other hand.

As with the SRRT test, the step sizes in the ‘diagonal’ condition were chosen to match the step size in the SRT test, in terms of the change in STI. This meant that, to make one step in the adaptive procedure, the signal-to-noise ratio was changed with √2 dB (i.e. $2^*\cos(\pi/4)$) , and, at the same time, the reverberation time was changed by a factor 1.28 (i.e. $\exp[\log(\sqrt{2})^*\cos(\pi/4)]$). These factors gives rise to a change in STI equivalent to 1.6 to 1.8 dB per step, for the relevant range (i.e. signal-to-noise ratios between -2 and +13 dB and reverberation times between 0.14 and 2.1 seconds). The step sizes of the adaptive procedures for conditions C and D were equal to the step sizes for the SRT (condition A) and the SRRT (condition B), respectively.

Measurement conditions C and D can be considered as reference conditions. Their inclusion makes it possible to assess the effect of the different adaptive procedures (horizontal, diagonal, or vertical) without changing the range of signal-to-noise ratios or reverberation times in which the measurement takes place.

B. Stimuli

The speech material used in the current experiment was the 'VU98' corpus (Versfeld et al., 2000), which consists of 39 lists of sentences of a male talker and the same amount of lists for a female talker. Only ten lists by the female speaker were used. The corpus was developed to enable efficient measurement of SRTs in noise. It was considered as being equivalent to the Plomp and Mimpen corpus (1979). However, the speaking style of the VU98 sentences is generally more informal, as expressed by the higher 'STI at the SRT' as discussed above.

Sentences were presented in stationary background noise, except for the SRRT test, in which no masker was present. Reverberation (if applicable) was introduced by convolving the auditory signals with synthetic impulse responses.* This approach makes it possible to define the reverberation time in a systematic way, excluding unwanted room effects which may be present when using real, recorded impulse responses. In terms of the modulation transfer function, this artificial reverberation is identical to purely exponentially decaying real reverberation of the same reverberation time. The impulse responses were created by subjecting white noise to a pure exponential decay. The slope of the decay and the length of the impulse response were determined by the desired reverberation time.

In all conditions, the long-term spectra of the speech and the masker were similar in shape. Reverberation (if present) was applied to the speech and to the stationary noise (if present), after which they were mixed according to the desired signal-to-noise ratio, based on their rms-levels. The spectrum of the resulting signal was adapted to reach octave masker levels equal to the middle of the dynamic range for each listener. The lower limit of the dynamic range was chosen to be the individual pure-tone threshold, while the upper limit was the uncomfortable loudness level (UCL), here chosen at 110 dB SPL. This approach is commonly used in our laboratory and is also applied for hearing-impaired listeners, as described below.

---

* To avoid onset artifacts of the reverberation, the convolution with the impulse response was always applied to a double sequence of the desired sentence in noise. Only the last half of the resulting signal was presented to the listener.

C. Participants

Five normal-hearing listeners participated in this experiment. Their ages were 26, 27, 46, 50 and 57 years. They reported no problems with their hearing or with speech reception, and were selected to have pure-tone hearing thresholds equal to or better than 10 dB HL at octave frequencies between 0.25 and 4.0 kHz.

D. General method and instrumentation

The experiment was run on a Dell personal computer, equipped with a Creative Labs Audigy external sound device and Beyer Dynamic DT48 headphones. Sound calibrations were performed with a Brüel & Kjær Artificial Ear (type 4152) and a Brüel & Kjær 2260 Observer conform ISO 389 (1991). All measurements were performed while listener and investigator were seated in a sound-insulated room, in a single half-an-hour session.

A test-retest design was followed, in which thresholds in each of the five conditions were measured twice for each participant. Test and retest outcomes were averaged. Confounding of measurement order and sentence lists with condition effects was avoided by counterbalancing the order of conditions across subjects, according to an five-by-five digram-balanced Latin square, while the order of the sentence lists was kept fixed. Measurements were conducted monaurally, using the participant's best ear, which was chosen according to personal preference in telephone conversation.

E. Data analysis

The measurement error in each of the listening conditions was compared with the measurement error in the commonly used SRT test, which is only about 1 dB (Plomp and Mimpen, 1979). Measurement accuracy was further evaluated by determining slopes of the psychometric curves. Logistic functions were fitted on the responses to the presented sentences (maximum likelihood) to determine, for each of the five measurement conditions, the proportion of correctly repeated sentences as a function of STI. For this purpose, the data were corrected for interindividual differences by shifting the individual data obtained for each test to the average STI for that condition (average over test, retest and participants).

		Condition				
		SRT	SRRT	horizontal	vertical	diagonal
SNR [dB]	mean	-3.9	--	+3.8	(2.8)	+3.2
	s.d.	1.1	--	1.2	--	0.3
T [s]	geom. mean	--	2.03	(0.89)	0.81	0.83
	geom. s.d.	--	1.16	--	1.17	1.05
STI	mean	0.371	0.344	0.385	0.378	0.385
	s.d.	0.038	0.030	0.021	0.022	0.012
STI	measurement error	0.030	0.017	0.019	0.026	0.031
	ΔSNR [dB]	0.89	0.49	0.56	0.78	0.93
	slope [%/dB]	28	25	31	27	18

Table 6-I Means and standard deviations of speech thresholds for normal-hearing listeners, for five conditions in which either the noise (SRT / horizontal), or the reverberation (SRRT / vertical) or both (diagonal) were adaptively varied to determine the threshold. Also shown are the corresponding STI-values with their standard deviations, the measurement errors, and the slopes of the fitted psychometric functions, transferred to easily understandable units (%/dB, see text). Measurement error in the observed STI-value (average of test and retest) was defined as the standard deviation of the test-retest difference, divided by √2.

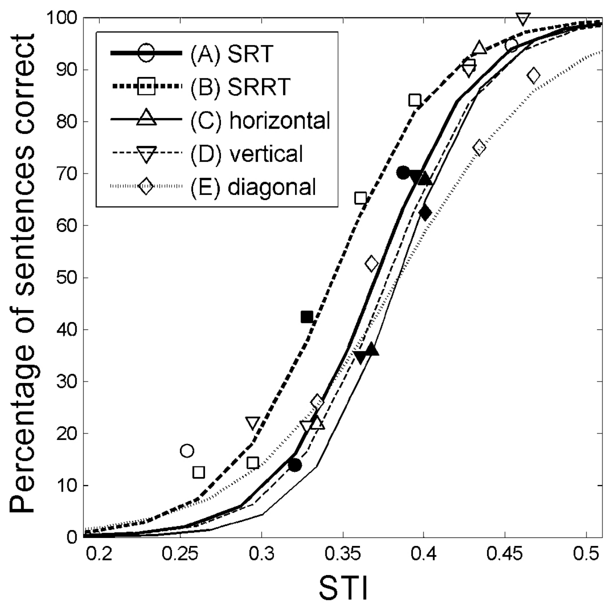

Figure 6-2 Estimated psychometric functions, based on a maximum likelihood fit, for sentences in noise and / or reverberation for normal-hearing listeners. Measurement conditions A to E as in Figure 1. Open symbols are based on at least five observations, closed symbols are based on more than thirty observations.

Finally, it was investigated whether the STI, averaged over conditions and participants, is an adequate single value to summarize the performance of normal-hearing listeners in each of the five measurement conditions. This was done by performing a repeated-measures Analysis of Variance (ANOVA), with 'condition' as a within-subjects effect.

F. Results

In Table 6-I, the means and standard deviations of the obtained thresholds are displayed, plus the means and standard deviations of the corresponding STI-values (averaged over test and retest). In all conditions, standard deviations are comparable to the standard deviation in the SRT test. This indicates that the range of the obtained STI-values is similar for all conditions. Also displayed are the measurement errors, which, for an easier evaluation, are also expressed in ΔSNR, representing the change in speech-to-noise ratio (without reverberation) that would be needed to give rise to the same change in STI. The measurement errors in all conditions are similar, or even better than, the measurement error of the SRT test. This indicates that, for normal-hearing listeners, the threshold for sentence intelligibility in each of the measurement conditions can be consistently determined.

The obtained psychometric functions for all five measurement conditions are shown in Fig. 6-2. Each curve is based on 130 responses, i.e. five listeners times two tests times thirteen presentations per test. The slopes of these functions at the steepest part of the curve (around 50% intelligibility) are shown in Table 6-I. The original slopes (unit: %/STI) were transferred to their current values (%/dB) by determining the change in speech-to-noise ratio (without reverberation) that would be needed to give rise to the same change in intellibility. A STI-change of 0.033 in STI is equivalent to a change of 1 dB SNR. The resulting slopes are comparable to the slope of about 20%/dB, as obtained for the SRT test by Plomp and Mimpen (1979).

Results of an Analysis of Variance (ANOVA) show that the effect of 'condition' was not significant ($p = 0.07$). Post-hoc comparisons (contrasts) were made between the STI in the individual conditions and the 'STI at the SRT', which is regarded as 'golden standard' here. None of these contrasts was significant ($p > 0.20$), which shows that the STI does not systematically vary across conditions.

G. Conclusion

It can be concluded that the threshold for sentence intelligibility can be consistently determined in all measurement conditions for normal-hearing listeners. Moreover, the applied adaptive procedure (horizontal, vertical, or diagonal) does not systematically affect the resulting 'STI at the threshold'. This finding makes the Speech Reception Reverberation Threshold (SRRT; vertical condition) an adequate test to examine the effect of solely reverberation on sentence intelligibilty, without the presence of noise.

III. EXPERIMENT 2: HEARING-IMPAIRED LISTENERS

A. Method and stimuli

Thresholds for sentence intelligibility were measured in three different conditions. These conditions were the same as applied in Experiment 1 (see Fig. 6-1). Sentences were presented in background noise only (condition A), in reverberation only (condition B) or in a combination of noise and reverberation (condition E). The earlier performed reference conditions C and D were not included in this experiment, since the results for normal-hearing listeners showed that the 'diagonal' measurement procedure adequately assesses the combined effect of reverberation and noise on speech intelligibility.

The auditory stimuli in Experiment 2 were also the same as used in Experiment 1. Tests were performed monaurally: depending on whether only one or both ears of the participant were included, six or twelve lists of the 'VU98' corpus were used in the measurements. Again, the long-term spectra of the speech and the masker were similar in shape for all conditions, and the spectrum of the stimuli was adapted to reach octave masker levels equal to the middle of the dynamic range for each listener. This approach assures optimal audibility for all listeners, thus minimizing effects due to differences in individual hearing thresholds.

B. Participants

Participants were patients of the audiology department of the VU University Medical Center. They were asked to participate in the experiment directly after a visit to the center for a clinical evaluation of their hearing loss. The pure-tone audiogram and CVC-word intelligibility scores (in quiet) of both ears were known from this clinical evaluation. Individual ears were included in the experiment when the average pure-tone threshold for the frequencies 500, 1000 and 2000 Hz was larger than 15 dB HL, i.e. when hearing was less than optimal. Moreover, a maximum CVC-word score of a least 70% was necessary to be included in the experiment. When both the listener's ears fitted these restrictions, both ears were measured and included separately.

A total of 61 ears from 36 participants were measured in the current experiment. Based on the audiogram and medical and family anamnesis, they were divided in four groups. A relatively large part of the included ears had a hearing loss diagnosed as presbycusic. The slope of the audiogram was used to divide these ears in two different groups. Slope was defined as the average pure-tone threshold at 2000 and 4000 Hz, minus the average threshold at 125 and 250 Hz. Group A consists of 27 presbycusic ears, from 16 presbycusic participants, with a slope larger than 25 dB. Ears included in this first group show hearing losses equivalent to those of the listeners as measured by Duquesnoy and Plomp (1980). Group B includes 11 ears, from 6 presbycusic participants, with a slope smaller than 25 dB. Group C consists of 8 ears, from 5 participants diagnosed with congenital hearing-impairement. Group D consists of 15 ears, from 9 participants with other types of hearing-impairment (due to medication, otosclerosis, chronic otitis in youth, or unknown cause). The hearing loss in all ears was sensorineural in nature, except for three ears in group D (from two participants), in which air-bone gaps of around 30 dB were observed.

C. General method and instrumentation

The experiment was run using the same equipment as in Experiment 1. A test-retest design was followed; thresholds in each of the three conditions were measured twice

			PTA	slope	CVC	age	Pure-tone threshold [dB HL] at … Hz						
	N		[dB HL]	[dB]	[%]	[years]	125	250	500	1000	2000	4000	8000
Group A	27	mean	34.8	36.6	97.6	75.3	20.7	20.9	26.7	30.4	47.4	67.4	82.0
		s.d.	9.0	13.1	7.1	6.3	9.9	10.0	11.1	11.3	11.4	13.7	13.0
Group B	11	mean	41.7	11.6	99.2	77.0	35.9	37.7	41.8	41.8	41.4	55.5	70.9
		s.d.	6.2	10.5	3.9	9.4	10.2	11.0	7.8	6.8	8.4	9.3	10.7
Group C	8	mean	41.0	10.9	99.9	52.4	33.8	35.6	37.5	38.1	47.5	43.8	52.5
		s.d.	13.9	13.0	1.1	14.9	10.3	12.1	17.3	19.1	8.9	14.6	16.7
Group D	15	mean	46.6	22.7	96.5	58.2	35.7	35.3	42.3	47.0	50.3	66.0	83.7
		s.d.	11.8	33.7	9.8	18.3	18.3	18.1	15.5	15.8	19.4	24.1	18.8

Table 6-II Means and standard deviations of the pure-tone thresholds for four groups of hearing-impaired ears. Also displayed are each group's average Pure Tone Average (PTA; 500, 100 and 2000 Hz), audiogram slope, maximum CVC-word intelligibility score in quiet, and age.

for each participant. Test and retest outcomes were averaged. Confounding of measurement order and sentence lists with condition was avoided by randomly choosing the order of conditions, for each ear, while sentence and list order was kept fixed. Measurements were conducted monaurally, starting with the participant's best ear (fitting the inclusion criteria), which was chosen according to the audiogram, or, in case of a symmetrical loss, personal preference in telephone conversation.

D. Data analysis

The measurement error in each of the listening conditions was compared with the measurement error in the SRT test.

Subsequently, it was investigated whether the monaural performance of an individual hearing-impaired listener could be summarized by a single STI-value. This was done by performing a repeated-measures Analysis of Variance (ANOVA), with 'condition' as a within-ears effect and 'group' as the between-ears factor. The same ANOVA was repeated with the Pure Tone Average ('PTA'), maximum CVC-score ('CVC'), audiometric slope, and age included in the analysis as covariates. CVC-scores were arcsine-transformed before conducting any analyses.

Finally, multiple stepwise regression analyses (MSRs) were performed to identify covariates that best account for differences between the ears in the observed speech reception, as expressed in the 'STI at the threshold' for each condition.

E. Results

Table 6-II displays, for each group of ears, the mean and standard deviation of the pure-tone thresholds, plus the average PTA, audiometric slope, maximum CVC-word intelligibility score and age. The observed speech intelligibility thresholds, in each of the three measurement conditions, are shown in Table 6-III. The STI-values corresponding to the thresholds are shown in Table 6-IV. As in the normal-hearing case, the standard deviations for the measured thresholds are comparable to the standard deviation of the SRT test, sometimes even smaller. Moreover, it can be seen that, for each group of ears and for each condition, the measurement error is comparable to the measurement error of the SRT test. This indicates that the speech reception threshold in each of the measurement conditions could be consistently determined for all hearing-impaired listeners.

Results of the first repeated-measures ANOVA, without any covariates, show a significant effect of 'condition' ($p < 0.001$) and an interaction effect of

			Condition					
			SRT		SRRT		diagonal	
		unit	M	S	M	S	M	S
Group A	SNR	dB	-0.2	1.6	--	--	+7.0	1.6
	T	sec.	--	--	0.73	1.48	0.43	1.31
Group B	SNR	dB	-0.7	1.3	--	--	+7.1	1.3
	T	sec.	--	--	0.88	1.29	0.42	1.25
Group C	SNR	dB	-1.6	3.5	--	--	+5.6	2.5
	T	sec.	--	--	1.10	1.58	0.54	1.51
Group D	SNR	dB	+0.3	2.9	--	--	+6.8	2.4
	T	sec.	--	--	0.80	1.70	0.44	1.48

Table 6-III Means (M) and standard deviations (S) of speech thresholds for four groups of hearing-impaired ears, for three conditions in which either the noise (SRT), the reverberation (SRRT), or both (diagonal) were adaptively varied to determine the threshold. Averages and standard deviations of reverberation times (T) were calculated geometrically.

'condition*group' ($p < 0.001$). When the four covariates (PTA, audiometric slope, CVC-score and age) are included in the analysis, the effect of 'condition' is no longer significant ($p > 0.20$), while the interaction effect of 'condition*group' still is significant ($p = 0.02$). Moreover, a significant interaction effect is observed for 'condition*age' ($p = 0.02$). When the covariates are included in the ANOVA, none of the differences in STI-value between individual conditions and the 'STI at the SRT' was significant in post-hoc comparisons (contrasts).

These results indicate that the original effect of 'condition' is mediated by the covariates, or, vice-versa, that the effect of the covariates on speech intelligibility performance changes with 'condition'. To further investigate this suggestion, Multiple Stepwise Regression Analyses (MSRs) were performed to identify which covariates contribute most to explaining variance between the ears in the observed STI-values, for each of the three measurement conditions. Results, as displayed in

		Condition		
	STI	SRT	SRRT	diagonal
Group A	mean	0.495	0.537	0.535
	s.d.	0.054	0.074	0.062
	measurement error	0.033	0.033	0.037
Group B	mean	0.477	0.502	0.539
	s.d.	0.043	0.052	0.053
	measurement error	0.035	0.037	0.039
Group C	mean	0.447	0.459	0.482
	s.d.	0.116	0.091	0.099
	measurement error	0.048	0.027	0.028
Group D	mean	0.508	0.518	0.526
	s.d.	0.097	0.102	0.092
	measurement error	0.038	0.035	0.030

Table 6-IV Averages and standard deviations of the STI-values corresponding to the thresholds in Table 6-III, for four groups of hearing-impaired ears. Measurement error in the observed STI-value (average of test and retest) was defined as the standard deviation of the test-retest difference, divided by √2.

	Condition					
	SRT		SRRT		diagonal	
	predictor	R^2	predictor	R^2	predictor	R^2
Step 1	CVC	.265	CVC	.297	CVC	.238
Step 2	--	--	Age	.395	Age	.330
Step 3	--	--	Slope	.429	--	--

Table 6-V Successive contributors to explaining the variance in STI-values between ears, for each of the three measurement conditions. Shown is the proportion of variance, corrected for the available degrees of freedom, that the relevant covariates cumulatively account for when included in the model. All shown models have a significance $p<0.001$.

Table 6-V, show that the maximum word intelligibility score (CVC) accounts for a large part of the variance in all three conditions. Age accounts for an additional amount of variance in the two conditions where reverberation was included. No other independent variables significantly contributed, except in the SRRT-condition, where audiogram slope explains a small, additional amount of variance in speech intelligibility performance. When, rather artificially, the audiogram (PTA) is forced in the regression model as a first step instead of the CVC-score, the effect of age is still significant in the conditions where reverberation is present.

In summary, these findings show that the observed 'STI at the threshold' changes significantly between measurement conditions. As the MSRs seem to confirm, this effect is mediated by the effect of 'age' on speech intelligibility performance, which appears to change over conditions. When this effect is accounted for by including the covariates in the ANOVA, no effect of 'condition' remains.

Scatter diagrams comparing the STIs in the three conditions are shown in Fig. 6.3. The standard deviations in STI of individual points from the diagonal are 0.043, 0.031, and 0.032 for the three panels, respectively. These deviations, although significant, are equivalent to a change in SNR smaller than 1.3 dB, which is similar to the measurement error. This means that, in clinical practice, an individual's monaural STI is still a good estimate for speech intelligibility performance, regardless of whether the speech is presented in background noise, in reverberation or in a combination of both.

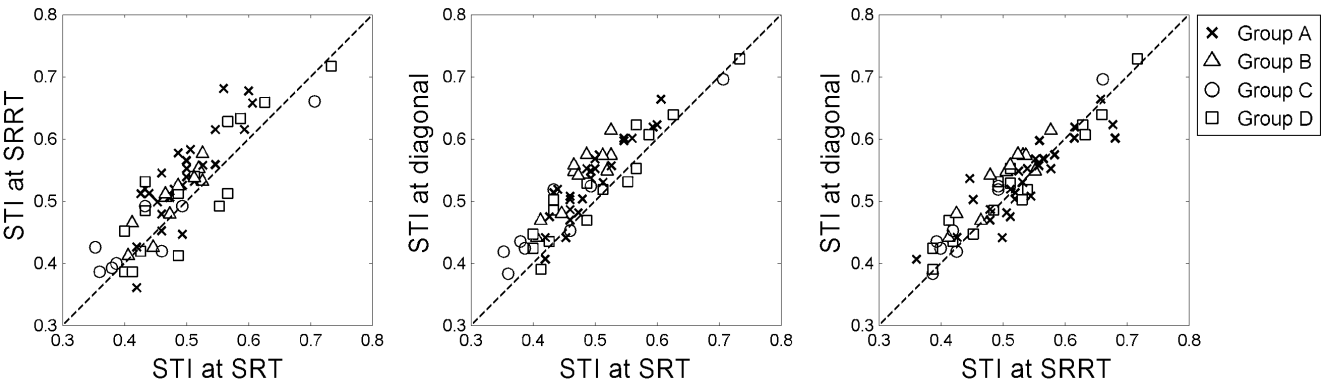

Figure 6-3 Scatter diagrams comparing the observed STIs in each of the three conditions. The diagonal lines represent where the values along the horizontal and vertical axis are equal. The standard deviations in STI of individual points from the diagonal are 0.043, 0.031, and 0.032 for the three panels, respectively.

F. Discussion

After including the covariates in the ANOVA, no relevant main effects were obtained in the analysis discussed above. A significant interaction was obtained, however, between condition and age. This interaction can be understood when considering the results from the MSRs, which show that age accounts for a significant part of the variance between ears in sentence reception when reverberation is present, but not when speech is presented in noise only. These differences in the contribution of age may be explained by a possible increased importance of cognitive processing in reverberant conditions, compared to non-reverberant conditions. Speech reception in stationary noise appears to be governed by auditory factors, like the audiogram (Van Rooij and Plomp, 1992; Smoorenburg, 1992; George et al., 2007) or CVC-scores, which explain a large part of variance of speech intelligibility performance in the current study (see Table 6-V). In reverberant conditions, age also plays a role, consistent with results by Helfer and Wilber (1990) and suggestions by Nabelek and Dagenais (1986). Kjellberg (2004) suggested that this age-effect is not only related to increased hearing loss with age, but is also related to the cognitive work load in listening situations with speech and reverberation. Similarly, Pichora-Fuller et al. (1995) suggest that the importance of (age-related) cognitive processing increases when listening situations are more complex. In addition, age-related differences in auditory temporal processing may play a role (Gordon-Salant and Fitzgibbons, 1999; Humes et al., 2007).

In addition, the ANOVA showed a significant interaction between measurement condition and group. This indicates that an effect of condition may be found within a specific group, even when there is no overall effect of condition. To further investigate this suggestion, the ANOVA was repeated within each group, but no significant effects of condition were observed ($p > 0.20$ in all cases). The fact that a significant interaction effect is nevertheless obtained, may be explained by the differences in age between the groups. Thus, the 'condition*group' interaction may be indirectly due to the 'condition*age' interaction, which was discussed above.

IV. FINAL CONCLUSION

In conclusion, the current findings indicate that the effect of measurement condition on STI is systematic and appears to be mediated by age. However, the differences between measurement conditions are comparable to the measurement error and can thus be considered fairly small. This means that the STI is still a convenient, single

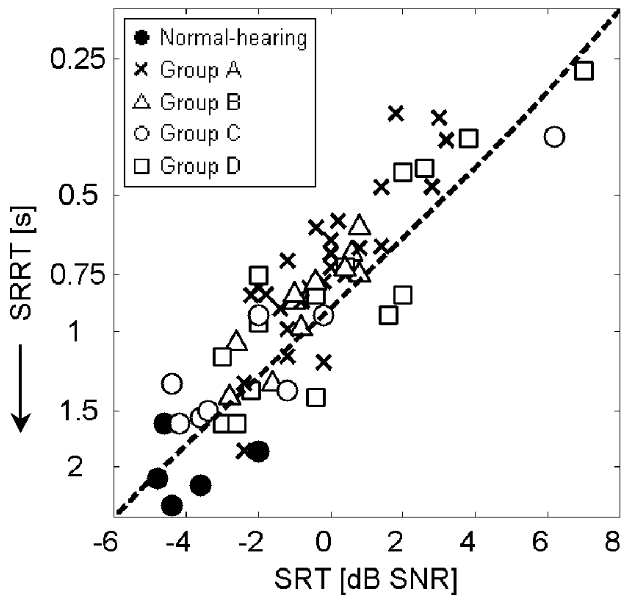

Figure 6-4 Scatter diagram of the Speech Reception Reverberation Threshold (SRRT) and the Speech Reception Threshold (SRT) for normal (solid circles) and hearing-impaired ears. The diagonal line represents where the STI-values along the horizontal and vertical axis are equal.

number to quantify speech recognition performance of a individual hearing-impaired listener, regardless of whether the sentences are presented in noise or in reverberation. The nature of the hearing loss (presbycusic, congenital, or otherwise) does not appear to affect this result.

The results of the current two experiments are summarized in Fig. 6-4, which gives the scatter plot of the individual scores for the SRRT and the SRT tests, for both normal-hearing and hearing-impaired listeners. A larger individual STI-value indicates a larger hearing loss for speech-in-noise, as described by the model by Plomp (1978). Since differences between the conditions are fairly small, a hearing loss for speech-in-noise may also be regarded as a hearing loss for speech-in-reverberation, as the scatter plot underlines. Nevertheless, age contributes to explaining differences in speech reception in reverberant conditions, while it does not play a role in accounting for variance in speech reception in noise only. Determining the effect of reverberation on speech intelligibility directly, using the Speech Reception Reverberation Threshold, may be helpful in understanding the possible role of cognitive processing in reverberant listening conditions.

V. ACKNOWLEDGEMENT

This research was supported by the Heinsius-Houbolt Foundation, The Netherlands.

7 General discussion and conclusion

I. GENERAL DISCUSSION AND CONCLUSION

A. Evaluation of goals

This thesis has investigated factors that affect speech reception for normal-hearing and hearing-impaired listeners. Two specific goals were stated in the Introduction, which will be evaluated here.

Our first goal was to investigate why hearing-impaired listeners benefit less from the relatively silent periods or 'gaps' in fluctuating noise than normal hearing listeners. Based on the results, it can be concluded (Chapter 2) that mainly deteriorated temporal resolution is related to the observed reduced masking release. In a subsequent experiment, the effect of temporal resolution was substituted by an effect of the audiogram. Auditory factors thus appear to be responsible for the main part of the variance in speech-in-noise reception. In addition, non-auditory factors contribute significantly to explaining inter-individual differences in speech intelligibility. Particulary measures of working memory and perceptual closure appear to be relevant for speech recognition, in line with expectations from literature (Humes et al., 2007; Zekveld et al., 2007a). Applying non-auditory measurements in a clinical setting may give rise to new insights concerning the rehabilitation of hearing-impaired listeners.

Our second goal was to investigate whether the relative contributions of noise and reverberation to disturbing speech intelligibility depend on the type of hearing-impairment. The current results show that this is not the case. This means that the STI-method of assessing reverberation and noise in terms of modulation transfer functions can be readily applied to hearing-impaired listeners with various types of hearing loss. The observed contribution of age to variance in the reception of reverberant speech again underlines the relevance of non-auditory factors for understanding sentences in complex listening situations.

B. Hearing loss and speech reception

In all experiments described in this thesis, we have tried to minimize the effect of hearing threshold elevation by optimizing audibility, for each individual listener. All auditory signals were adapted to reach octave masker levels equal to the middle of the dynamic range for each listener. The lower limit of the dynamic range was chosen to be the individual pure-tone threshold, while the upper limit was the uncomfortable loudness level (or 110 dB SPL). Unfortunately, this procedure gives rise to inter-individual spectrum and level differences. Using the SRT to quantify the subjects' ability to perceive speech in noise does not take these differences into account. In addition, a direct consequence of the applied optimization method is that presentation level and hearing threshold become confounded, and, as a consequence, their effects on speech reception cannot be fully distinguished. Thus, minimizing the direct effect of hearing loss by optimizing audibility *increases* possible indirect effects of hearing loss due to level differences.

These problems have been assessed by applying the Speech Intelligibility Index or SII, which is able to handle inter-subject differences in audiogram, spectrum, and level. Results show, however, that even when the SII model is applied, the pure-tone hearing loss is still the best predictor of speech reception in both stationary and non-stationary noise. It is argued that this can be explained by the large interdependency between hearing loss and supra-threshold deficits: the development of hearing loss generally accompanies deterioration of supra-threshold processing, and vice-versa.

An alternative method to distinguish between threshold-related and supra-threshold problems in speech reception has been applied in Chapter 2 of this thesis. The experiment did not only include normal-hearing and hearing-impaired listeners, but also a third group of subjects with a simulated hearing loss (by presenting an additional noise to artificially raise pure-tone hearing thresholds). Comparing the three groups makes it possible to adequately distinguish between threshold-related and supra-threshold related deficits. Nevertheless, like before, hearing loss was significantly related to the variance in masking release, although it played no significant role anymore when temporal resolution was included in the regression model.

It must be concluded that hearing loss always appears to be an important factor in the results, affecting speech intelligibility either directly (via audibility) or indirectly (via level effects or via the relation with supra-threshold processing). Thus, even though supra-threshold processing appears to underlie speech-in-noise reception, consistent with the model by Plomp (1978), the audiogram appears to be a significant predictor of speech-in-noise performance for hearing-impaired listeners.

C. Temporal factors and speech reception

Temporal acuity has been shown to be the main factor affecting the benefit for speech reception from masker fluctuations. Auditory spectral acuity did not appear to be play a role in the results. Apparently, listeners with broadened spectral filters still have sufficient spectral resolution to resolve the spectral cues important for speech intelligibility, consistent with the results by Shannon et al. (1995). These findings suggest that mainly temporal factors are relevant for auditory processing and the reception of speech in everyday listening situations.

Three different measures have been used in this thesis which are related specifically to temporal processing in the auditory system: i) temporal acuity (ΔT) measured by tone sweeps in temporal noise grids; ii) the Speech Reception Timewidth Threshold (SRTT), measured with complementary chopped speech and noise; iii) the Speech Reception Reverberation Threshold (SRRT), assessing solely the effect of reverberation in quiet. Future research might investigate the role of temporal processing in speech reception further by evaluating the mutual dependencies among these measures, and their correlations with other auditory or non-auditory abilities. Specifically in the light of a possible specialized speech pathway in the human cortex, as discussed below, it is important to further evaluate the relations between temporal acuity as measured with tone sweeps and the temporal resolution relevant for understanding speech in noise or reverberation.

D. Further understanding speech reception

It has been shown that a substantial portion of the variance in speech-in-noise reception cannot be explained by the auditory and nonauditory variables measured as in the current thesis. It has been argued that one explanation for the amount of unexplained variance may be the limited selection of auditory and nonauditory tests, even though specifically those tests were selected that were found in literature to be related to speech reception in noise. Another explanation is that no tests were included that measure cognitive or other supra-cochlear factors that are specifically auditory in nature. To further investigate this possibility, future research may focus on combining measurements of modality-aspecific, cognitive and verbal processing with modality-specific, auditory tests of peripheral and non-peripheral processing. This approach, combining tests at three levels of speech-in-noise processing (auditory/peripheral, auditory/non-peripheral, non-auditory/non-peripheral) may enable identifying factors that account for the remaining unexplained variance. Moreover, it was mentioned that supra-threshold auditory processing was only measured in the frequency region around 1 kHz in the current experiments. Performing these measurements in a broader frequency range might also improve the amount of explained variance.

There may be yet another explanation for the relatively large portion of unexplained variance in speech reception. When accounting for variance with auditory measures, it is assumed that the tones and noises, as applied in the experiments, are processed in the human auditory system in a way similar to speech. Likewise, it is assumed that the same cognitive capacity is available for speech reception as for the performance of other cognitive tasks. However, results by Whalen et al. (2006) suggest that the human cortex includes a specialized speech system, which bypasses part of the processing in the primary auditory cortex. Scott et al. (2000) also identified a separate brain pathway for understanding intelligible speech. Thus, it appears that the brain has a separate unit for the processing of speech, that may use other auditory and cognitive capacities than those that are used when measuring non-speechlike variables (temporal resolution, working memory). Therefore, future research on the relation of auditory and non-auditory factors with speech-in-noise reception should at least consider measuring these factors with speech-like stimuli. This perspective underlines the relevance of the developed Text Reception Threshold as a tool to determine the non-auditory factors in speech reception.

E. Limitations and improvements

It should be noted that our results only apply to listeners with mild to moderate hearing-impairment. For most of the current experiments, the listener's audiometric threshold were selected to be no larger than 60 dB HL on average. In Chapter 6, a CVC-intelligibility score (in quiet) of at least 70% was required to be included as a participant. For more severly hearing-impaired listeners, speech reception may depend even more on the reduced sensitivity and acuity of auditory processing. Alternatively, however, it may be argued that a larger hearing loss might increase the complexity of the listening situation and the cognitive load for a listener, thus increasing the relative importance of non-auditory factors (see Pichora-Fuller et al., 1995). Which of these factors prevails can, for the moment, only remain subject to speculation.

Moreover, it should be mentioned that our experiments concerning the relevance of non-auditory factors for speech intelligibility in fluctuating noise were performed using block-modulated maskers with a large modulation depth. These maskers were chosen to assess masking release for speech intelligibility as adequately as possible. Fluctuating backgrounds in everyday life commonly contain a broader modulation spectrum and less extreme modulations. Therefore, the effects related to 'listening in the gaps' may be expected to be less pronounced than obtained in our experimental setting.

Finally, effects related to binaural listening were not taken into account in the current study. Using the differences in signal level and arrival time between the two ears, listeners can obtain a binaural or spatial release from masking (the 'cocktail party effect'). Bronkhorst and Plomp (1992) showed that this binaural effect is similar in size for normal-hearing and hearing-impaired listeners. However, more recent studies (Noble et al., 1997; Gatehouse and Noble, 2004; Goverts, 2004) indicate that hearing loss can adversely affect spatial release from masking. Including binaural effects in models of speech intelligibility like the STI (see Van Wijngaarden and Drullman, 2006) may contribute to further understanding the differences between normal-hearing and hearing-impaired listeners when it comes to speech intelligibility in daily life.

F. Clinical relevance

In clinical audiology, the importance of adequate quantification of the problems that an individual listener experiences in understanding speech can not be overestimated. The audiogram and maximum intelligibility scores for CVC-words in quiet, measured as part of the standard clinical procedure in the Netherlands, are often insufficient to account for the large variation between hearing-impaired listeners in speech intelligibility. The Speech Reception Threshold (SRT) in stationary noise appears to be a reasonable estimate of the supra-threshold deficits of an individual listener (Stephens, 1976; Plomp, 1978). However, it is unclear how this SRT relates to the problems that a listener faces in everyday listening situations, in which speech information is commonly disturbed by fluctuating background noise (e.g. competing talkers) and reverberation.

It has been suggested (Pichora-Fuller, 1995) that as the complexity of a listening situation increases, the importance of non-auditory factors for speech reception increases likewise. The current results confirm this suggestion. They show that, besides auditory factors, non-auditory factors are relevant for understanding sentences in non-stationary noise, specifically working memory capacity (SWM) and perceptual closure skills (TRT; see Chapter 4). The obtained relevance of age in understanding speech in reverberant conditions also seems to corroborate this suggestion (see Chapter 6). In less complicated conditions with solely stationary noise as a masker, non-auditory factors do not contribute significantly to sentence intelligibility, consistent with results by earlier studies (Van Rooij and Plomp, 1992).

Therefore, the diagnostics of hearing-impairment in listeners who complain of not being able to understand speech, should start with measuring thresholds for sentence intelligibility in more complex listening situations, that is, in fluctuating noise or in reverberation. Measuring the SRT in fluctuating noise offers a better measure for everyday speech communication (see Kramer, 1996). Moreover, the SRT in fluctuating noise appears to be related to supra-threshold deficits in hearing (see Chapter 2), and also appears to take into account the contribution of non-auditory processing to speech perception.

The combined measurement of SRT in fluctuating noise and non-auditory factors (SWM or TRT) makes it possible to estimate the relative contribution of auditory and non-auditory factors to speech recognition in noise (Chapter 3 and 4). This may enable the clinical audiologist to determine in part the origin (auditory or non-auditory) of deteriorated speech reception. The expectations concerning the benefit from auditory rehabilitation can then be adapted likewise (see Kricos, 2006). In addition, the current results suggest that improving non-auditory processing by training might contribute to speech intelligibility performance, and thus may be relevant for auditory rehabilitation in the future.

To conclude, it should be stressed that understanding speech in everyday life is still very difficult for hearing-impaired listeners. Considering the large prevalence of hearing-impairment, efforts should be made to improve listening situations in cases where effective communication is of vital importance. Furthermore, it is suggested that the measurement of speech intelligibility thresholds in complex listening situations (reverberation and fluctuating noise) is incorporated in standard clinical procedures for hearing-impaired listeners. Our findings show that the clinical measurement of non-auditory factors (TRT and SWM) makes it possible to determine, in part, the source of the experienced problem in speech reception (non-auditory or auditory). Therefore, non-auditory factors should be taken into account when clinically assessing the challenges of an individual hearing-impaired listener in daily life.

References

REFERENCES

Allen, M.J., and Yen, W.M. (**1979**). *Introduction to measurement theory.* Brooks-Cole, Montery, CA.

Amitay, S., Ahissar, M., and Nelken, I. (**2002**). "Auditory processing deficits in reading disabled adults," J. Assoc. Res. Otolaryngol., **3**, 302-320.

ANSI (**1997**). ANSI S3.5-1997, "American national standard methods for the calculation of the Speech Intelligibility Index," American National Standards Institute, New York.

Apoux, F., and Bacon, S.P. (**2004**). "Relative importance of temporal information in various frequency regions for consonant identification in quiet and in noise," J. Acoust. Soc. Am. **116**, 1671-1680.

Bacon, S.P., Opie, J.M., and Montoya, D.Y. (**1998**). "The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds," J. Speech Lang. Hear. Res. **41,** 549-563.

Baer, T., and Moore, B.C.J. (**1993**). "Effects of spectral smearing on the intelligibility of sentences in noise," J. Acoust. Soc. Am. **94**, 1229-1241.

Baer, T., and Moore, B.C.J. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277-2280.

Bilger, R.C., Nuetzel, J.M., Rabinowitz, W.M., and Rzeczkowski, C. (**1984**). "Standardization of a test of speech perception in noise," J. Speech Hear. Res. **27**, 32-48.

Boothroyd, A., Mulhearn, B., Gong, J., and Ostroff, J. (**1996**). "Effects of spectral smearing on phoneme and word recognition," J. Acoust. Soc. Am. **100**, 1807-1818.

Boothroyd, A., and Nittrouer, S. (**1988**). "Mathematical treatment of context effects in phoneme and word recognition," J. Acoust. Soc. Am. **84**, 101-114.

Bronkhorst, A.W., and Plomp, R. (**1992**). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," J. Acoust. Soc. Am. **92**, 3132-3139.

Celmer, R.D., and Bienvenue, G.R. (**1987**). "Critical bands in the perception of speech signals by normal and sensorineural hearing loss listeners," in *The Psychophysics of Speech Perception*, edited by M.E.H. Schouten (Nijhoff, Dordrecht).

Committee on Hearing, Bioacoustics, and Biomechanics (CHABA), Working Group on Speech Understanding (**1988**). "Speech understanding and aging," J. Acoust. Soc. Am. **83**, 859-895.

Davis, S.M., and McCroskey, R.L. (**1980**). "Auditory fusion in children," Child Dev. **51**, 75-80.

Divenyi, P.L., Stark, P.B., and Haupt, K.M. (**2005**). "Decline of speech understanding and auditory thresholds in the elderly," J. Acoust. Soc. Am. **118**, 1089-1100.

Drullman, R., Festen, J.M., and Plomp, R. (**1994**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053-1064.

Dubno, J.R., Ahlstrom, J.B., and Horwith, A.R. (**2000**). "Use of context by young and aged adults with normal hearing," J. Acoust. Soc. Am. **107**, 538-546.

Dubno, J.R., Horwitz, A.R., and Ahlstrom, J.B. (**2002**). "Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing," J. Acoust. Soc. Am. **111**, 2897-2907.

Dubno, J.R., Horwitz, A.R., and Ahlstrom, J.B. (**2003**). "Recovery from prior stimulation: masking of speech by interrupted noise for younger and older adults with normal-hearing," J. Acoust. Soc. Am. **113**, 2084-2094.

Dubno, J.R., and Schaefer, A.B. (**1992**). "Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners," J. Acoust. Soc. Am. **91**, 2110-2121.

Duifhuis, H (**1973**). "Consequences of peripheral frequency selectivity for nonsimultaneous masking," J. Acoust. Soc. Am. **54**, 1471-1488.

Duquesnoy, A.J. (**1983**). "The intelligibility of sentences in quiet and in noise in aged listeners," J. Acoust. Soc. Am. **74**, 1136-1144.

Duquesnoy, A.J., and Plomp, R. (**1980**). "Effect of reverberation and noise on the intelligibility of sentences in cases of presbyacusis," J. Acoust. Soc. Am. **68**, 537-544.

Eisenberg, L.S., Dirks, D.D., and Bell, T.S. (**1995**). "Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing," J. Speech Lang. Hear. Res. **38**, 222-233.

Elliott, L.L. (**1995**). "Verbal auditory closure and the Speech Perception in Noise (SPIN) test," J. Speech Hear. Res. **38**, 1363-1376.

Fabry, D.A., and Van Tasell, D.J. (**1986**). "Masked and filtered simulation of hearing loss: effects on consonant recognition," J. Speech Hear. Res. **29**, 170-178.

Festen, J.M. (**1993**). "Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by an interfering voice," J. Acoust. Soc. Am. **94**, 1295-1300.

Festen, J.M., and Plomp, R. (**1981**). "Relations between auditory functions in normal hearing," J. Acoust. Soc. Am. **70**, 356-369.

Festen, J.M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal-hearing," J. Acoust. Soc. Am. **88**, 1725-1736.

Fitzgibbons, P.J. (**1983**). "Temporal gap detection in noise as a function of frequency, bandwidth, and level," J. Acoust. Soc. Am. **74**, 67-72.

Fitzgibbons, P.J., and Gordon-Salant, S. (**1987**). "Temporal gap resolution in listeners with high-frequency sensorineural hearing loss," J. Acoust. Soc. Am. **81**, 133-137.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47-65.

Florentine, M., and Buus, S. (**1984**). "Temporal gap detection in sensorineural and simulated hearing impairments," J. Speech Hear. Res. **27**, 449-455.

French, N.R., and Steinberg, J.C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90-119.

Füllgrabe, C., Berthommier, F., and Lorenzi, C. (**2006**). "Masking release for consonant features in temporally fluctuating background noise," Hearing Res. **211**, 74-84.

Gatehouse, S., Naylor, G., and Elberling, C. (**2003**). "Benefits from hearing aids in relation to the interaction between the user and the environment," Int. J. Audiol. **42,** S77-85

Gatehouse, S., and Noble, W. (**2004**). "The Speech, Spatial and Qualities of Hearing scale (SSQ)," Int. J. Audiol. **43**, 85-99.

George, E.L.J., Festen, J.M., and Houtgast, T. (**2006**). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 2295-2311.

George, E.L.J., Zekveld, A.A., Kramer, S.E., Goverts, S.T., Festen, J.M., and Houtgast, T. (**2007**). "Auditory and nonauditory factors affecting speech reception in noise by older listeners," J. Acoust. Soc. Am. **121**, 2362-2375.

Gifford, R.H., and Bacon, S.P. (**2005**). "Psychophysical estimates of nonlinear cochlear processing in younger and older listeners," J. Acoust. Soc. Am. **118**, 3823-3833.

Glasberg, B.R., and Moore, B.C.J. (**1989**). "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech," Scand. Audiol. Suppl. **32**, 1-25.

Glasberg, B.R., and Moore, B.C.J. (**1992**). "Effects of envelope fluctuations on gap detection," Hearing Res. **64**, 81-92.

Glasberg, B.R., Moore, B.C.J., and Bacon, S.P. (**1987**). "Gap detection and masking in hearing-impaired and normal-hearing subjects," J. Acoust. Soc. Am. **81**, 1546-1556.

Goldstein, E. B. (**2002**). *Sensation and perception.* Wadsworth-Thomson, Pacific Grove, CA.

Gordon-Salant, S. (**1987**). "Age-related differences in speech recognition performance as a function of test format and paradigm," Ear Hear. **8**, 277-282.

Gordon-Salant, S., and Fitzgibbons, P.J. (**1997**). "Selected cognitive factors and speech recognition performance among young and elderly listeners," J. Speech Lang. Hear. Res. **40**, 423-431.

Gordon-Salant, S., and Fitzgibbons, P.J. (**1999**). "Profile of auditory temporal processing in older listeners," J. Speech Lang. Hear. Res. **42**, 300-311.

Gordon-Salant, S., and Fitzgibbons, P.J. (**2004**). "Effects of stimulus and noise rate variability on speech perception by younger and older adults," J. Acoust. Soc. Am. **115**, 1808-1817.

Goverts, S.T. (**2004**). *Assessment of spatial and binaural hearing in hearing-impaired listeners.* Doctoral thesis, VU University, Amsterdam, The Netherlands.

Grant, K.W., and Seitz, P.F. (**2000**). "The recognition of isolated words and words in sentences: individual variability in the use of sentence context," J. Acoust. Soc. Am. **107**, 1000-1011.

Grant, K.W., Walden, B.E., and Seitz, P.F. (**1998**). “Auditory-visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition, and auditory-visual integration,” J. Acoust. Soc. Am. **103**, 2677-2690.

Green, D.M., and Swets, J.A. (**1974**). *Signal detection theory and psychophysics.* Krieger, Huntington, NY.

Grimm, L.G., and Yarnold, P.R. (**1995**). *Reading and understanding multivariate statistics.* American Psychological Association, Washington, D.C.

Grose, J.H., Hall, J.W., and Buss, E. (**2001**). “Gap duration discrimination in listeners with cochlear hearing loss: effects of gap and masker duration, frequency separation, and mode of presentation,” J. Assoc. Res. Otolaryngol. **2**, 388-398.

Gustafsson, H.A., and Arlinger, S.D. (**1993**). “Masking of speech by amplitude-modulated noise,' J. Acoust. Soc. Am. **95**, 518-529.

Hall, J.W., Davis, A.C., Haggard, M.P., and Pillsbury, H.C. (**1988**). “Spectro-temporal analysis in normal-hearing and cochlear-impaired listeners,” J. Acoust. Soc. Am. **84**, 1325-1331.

Hall, J.W., Haggard, M.P., and Fernandes, M.A. (**1984**). “Detection in noise by spectro-temporal pattern analysis,” J. Acoust. Soc. Am. **76**, 50-56.

Hällgren, M. (**2005**). *Hearing and cognition in speech comprehension.* Doctoral thesis, Linköping University, Sweden.

Hawkins, J.E., and Stevens, S.S. (**1950**). “The masking of pure tones and of speech by white noise,” J. Acoust. Soc. Am. **22**, 6-13.

Healy, E.W., and Bacon, S.P. (**2006**). “Measuring the critical band for speech,” J. Acoust. Soc. Am. **119**, 1083-1091.

Helfer, K.S., and Wilber, L.A. (**1990**). “Hearing loss, aging, and speech perception in reverberation and noise,” J. Speech. Hear. Res. **33**, 149-155.

Hilkhuysen, G.L.M., Houtgast, T., and Lyzenga, J. (**2005**). “Estimating cochlear-filter shapes, temporal-window width and compression from tone-sweep detection in spectral and temporal noise gaps," J. Acoust. Soc. Am. **117,** 2598.

Hillenbrand, J.M., Clark, M.J., and Nearey, T.M. (**2001**). “Effects of consonant environment on vowel formant patterns,” J. Acoust. Soc. Am. **109**, 748-763.

Hogan, C.A., and Turner, C.W. (**1998**). “High-frequency audibility: benefits for hearing-impaired listeners,” J. Acoust. Soc. Am. **104**, 432-441.

Houtgast, T., and Steeneken, H.J.M. (**1984**). “A multi-lingual evaluation of the Rasti-method for estimating speech intelligibility in auditoria,” Acustica **54**, 185-199.

Houtgast, T., and Steeneken, H.J.M. (**1985**). “A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria,” J. Acoust. Soc. Am. **77**, 1069-1077.

Houtgast, T., Steeneken, H.J.M., and Plomp, R. (**1980**). “Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics,” Acustica **46**, 59-72.

Howard-Jones, P.A., and Rosen, S. (**1993**). "Uncomodulated glimpsing in "checkerboard" noise," J. Acoust. Soc. Am. **93**, 2915-2922.

Humes, L.E. (**2002**). "Factors underlying the speech-recognition performance of elderly hearing-aid wearers," J. Acoust. Soc. Am. **112,** 1112-1132.

Humes, L.E. (**2005**). "Do 'auditory processing' tests measure auditory processing in the elderly?," Ear Hear. **26**, 109-119.

Humes, L.E., Burk, M.H., Coughlin, M.P., Busey, T.A., and Strausser, L.E. (**2007**). "Auditory speech recognition and visual text recognition in younger and older adults: similarities and differences between modalities and the effects of presentation rate," J. Speech. Lang. Hear. Res. **50**, 282-303.

Humes, L.E., Dirks, D.D., Bell, T.S., and Kincaid, G.E. (**1987**). "Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers," J. Acoust. Soc. Am. **81**, 765-773.

Humes, L.E., Espinoza-Varas, B., and Watson, C.S. (**1988**). "Modeling sensorineural hearing loss. I. Model and retrospective evaluation," J. Acoust. Soc. Am. **83**, 188-202.

Hygge, S., Ronnberg, J., Larsby, B. and Arlinger, S. (1992). "Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds," J. Speech Hear. Res. **35**, 208-215.

ISO, International Organization for Standardization (**1991**). ISO 389:1991(E). *Acoustics – Standard reference zero for the calibration of pure-tone air conduction audiometers.* Available from the American National Standards Institute, New York.

Ishihara, S. (**1989**). *The series of plates designed as a test for colour deficiency.* Kanehara, Tokyo.

Jerger, J., Jerger, S., and Pirozzolo, F. (**1991**). "Correlational analysis of speech audiometric scores, hearing loss, age and cognitive abilities in the elderly," Ear Hear. **12**, 103-109.

Jesteadt, W., Bacon, S.P., and Lehman, J.R. (**1982**). "Forward masking as a function of frequency, masker level, and signal delay," J. Acoust. Soc. Am. **71**, 950-962.

Kalikow, D.N., Stevens, K.N., and Elliott, L.L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337-1351.

Kjellberg, A. (**2004**). "Effects of reverberation time on the cognitive load in speech communication: theoretical considerations," Noise Health **7**, 11-21.

Knutson, J.F., Schartz. H.A., Gantz, B.J. Tyler, R.S., Hinriches, J.V. and Tyler, R.S. (**1991**). "Psychological change following 18 months of cochlear implant use," Ann. Otol. Rhinol. Laryngol. **100**, 877-82.

Kramer, S.E., Kapteyn, T.S., Festen, J.M., and Tobi, H. (**1996**). "The relationships between self-reported hearing disability and measures of auditory disability," Audiol. (**35**), 277-287.

Kricos, P.B. (**2006**). "Audiologic management of older adults with hearing loss and compromised cognitive/pscyhoacoustic auditory processing capabilities," Trends Amplif. **10**, 1-28.

Kryter, K.R. (**1962**). "Methods for the calculation and use of the Articulation Index," J. Acoust. Soc. Am. **34**, 1689-1697.

Las, L., Stern, E.A., and Nelken, I. (**2005**). "Representation of tone in fluctuating maskers in the ascending auditory system," J. Neurosci. **25**, 1503-1513.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467-477.

Liberman, A.M., Harris, K.S., Hoffman, H.S.,, Griffith, B.C. (**1957**). "The discrimination of speech sounds within and across phoneme boundaries,"J. Exp. Psych. **54**, 358-368.

Lochner, J.P.A., and Burger, J.F. (**1961**). "Form of the loudness function in the presence of masking noise," J. Acoust. Soc. Am. **33**, 1705-1707.

Ludvigsen, C. (**1985**). "Relations among some psychoacoustic parameters in normal and cochlearly impaired listeners," J. Acoust. Soc. Am. **78**, 1271-1280.

Lunner, T. (**2003**). "Cognitive function in relation to hearing aid use," Int. J. Audiol. **42**, S49-58.

Miller, R.G. (**1981**). *Simultaneous Statistical Inference.* Springer, New York.

Moore, B.C.J. (**2003**). *An introduction to the psychology of hearing; fifth edition.* Academic Press, San Diego, CA.

Moore, B.C.J., Peters, R.W., and Glasberg, B.R. (**1993**). "Detection of temporal gaps in sinusoids: effects of frequency and level," J. Acoust. Soc. Am. **93**, 1563-1570.

Moore, B.C.J., Vickers, D.A., Plack, C.J., and Oxenham, A.J. (**1999**). "Inter-relationship between different psycho-acoustic measures assumed to be related to the cochlear active mechanism," J. Acoust. Soc. Am. **106**, 2761-2778.

Nabelek, A.K., and Dagenais, P.A. (**1986**). "Vowel errors in noise and in reverberation by hearing-impaired listeners," J. Acoust. Soc. Am. **80**, 741-748.

Nabelek, A.K., and Mason, D. (**1981**). "Effects of noise and reverberation on binaural and monaural word identification by subjects with various audiograms," J. Speech Hear. Res. **24**, 375-383.

Nabelek, A.K., and Robinson, P.K. (**1982**). "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. 71, 1242-1248.

Nelson, P.B., and Jin, S.H. (**2004**). "Factors affecting speech understanding in gated interference: cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. 115, 2286-2294.

Nelson, P.B., Jin, S.H., Carney, A.E. and Nelson, D.A., (**2003**). "Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 962-968.

Nittrouer, S., and Boothroyd, A. (**1990**). "Context effects in phoneme and word recognition by young children and older adults," J. Acoust. Soc. Am. **87**, 2705-2715.

Noble, W., Byrne, D., and Ter-Horst, K. (**1997**). “Auditory localization, detection of spatial separateness, and speech hearing in noise by hearing-impaired listeners,” J. Acoust. Soc. Am. **102**, 2343-2352.

Noordhoek, I.M., Houtgast, T., and Festen, J.M. (**1999**). “Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. I. Normal-hearing listeners,” J. Acoust. Soc. Am. **105**, 2895-2902.

Noordhoek, I.M., Houtgast, T., and Festen, J.M. (**2000**). “Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. II. Hearing-impaired listeners,” J. Acoust. Soc. Am. **107**, 1685-1696.

Noordhoek, I.M., Houtgast, T., and Festen, J.M. (**2001**). “Relations between intelligibility of narrow-band speech and auditory functions, both in the 1-kHz frequency region,” J. Acoust. Soc. Am. **109**, 1197-1212.

Nunnally, J.C. (**1967**). *Psychometric Theory.* McGraw-Hill, New York.

Owen, A.M., Downes, J.J., Sahakian, B.J., Polkey, C.E., and Robbins, T.W. (**1999**). “Planning and spatial working memory following frontal lobe lesions in man,” Neuropsychologia*,* **28**, 1021-1034.

Oxenham, A.J., and Bacon, S.P (**2003**). “Cochlear compression: perceptual measures and implications for normal and impaired hearing,” Ear Hear. **24**, 352-366.

Patterson, R.D., Nimmo-Smith, I., Weber, D.L., and Milroy, R. (**1982**). “The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold,” J. Acoust. Soc. Am. **72**, 1788-1803.

Pavlovic, C.V. (**1987**). “Derivation of primary parameters and procedures for use in speech intelligibility predictions,” J. Acoust. Soc. Am. **82**, 413-422.

Peters, R.W., Moore, B.C.J., and Baer, T. (**1998**). “Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people,” J. Acoust. Soc. Am. **103**, 577-587.

Pichora-Fuller, M.K., Schneider, B.A., and Daneman, M. (**1995**). “How young and old adults listen to and remember speech in noise,” J. Acoust. Soc. Am. **97**, 593-608.

Pichora-Fuller, M.K., and Souza, P.E. (**2003**). “Effects of aging on auditory processing of speech,” Int. J. Aud. **42**, 2S11-2S16.

Plomp, R. (**1978**). “Auditory handicap of hearing impairment and the limited benefit of hearing aids,” J. Acoust. Soc. Am. **63**, 533-549.

Plomp, R., and Mimpen, A.M. (**1979**). “Improving the reliability of testing the speech reception threshold for sentences,” Audiol. **18**, 43-52.

Rhebergen, K.S., and Versfeld, N.J. (**2005**). “A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners,” J. Acoust. Soc. Am. **117**, 2181-2192.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., McInnes, L. and Rabbitt, P. M. (**1994**). "The Cambridge Neuropsychological Test Automated Battery CANTAB; a factor analytical study in a large number of normal elderly volunteers," Dementia **5**, 266-281.

Sato, H., Sato, H., and Morimoto, M. (**2007**). "Effects of aging on word intelligibility and listening difficulty in various reverberant fields," J. Acoust. Soc. Am. **121**, 2915-2922.

Scott, S.K., Blank, C.C., Rosen, S., and Wise, R.J.S. (**2000**). "Identification of a pathway for intelligible speech in the left temporal lobe," Brain **123**, 2400-2406.

Shannon, R.V., Zeng, F.-G., Kamath, V., Wygonski, J., Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303-304.

Shaw, E.A.G., and Teranishi, R. (**1968**). "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," J. Acoust. Soc. Am. **44**, 240-249.

Smoorenburg, G.F. (**1992**). "Speech reception in quiet and noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram," J. Acoust. Soc. Am. **91**, 421-437.

Snell, K.B. (**1997**). "Age-related changes in temporal gap detection," J. Acoust. Soc. Am. **101**, 2214-2220.

Snell, K.B., Mapes, F.M., Hickman, E.D., and Frisina, D.R. (**2002**). "Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity," J. Acoust. Soc. Am. **112**, 720-727.

Snodgrass, J.G., and Kinjo, H. (**1998**). "On the generality of the perceptual closure effect," J. Exp. Psychol. Learn. Mem. Cogn. **24**, 645-658.

Sommers, M.S., and Humes, L.E. (**1993a**). "Auditory filter shapes in normal-hearing, noise-masked normal and elderly listeners," J. Acoust. Soc. Am. **93**, 2903-2914.

Sommers, M.S., and Humes, L.E. (**1993b**). "Erratum: Auditory filter shapes in normal-hearing, noise-masked normal and elderly listeners [J. Acoust. Soc. Am. **93,** 2903-2914 (1993)]," J. Acoust. Soc. Am. **94**, 2449-2450.

Souza, P.E., and Turner, C.W. (**1994**). "Masking of speech in young and elderly listeners with hearing loss," J. Speech Hear. Res. **37**, 655-661.

Steeneken, H.J.M., and Houtgast, T. (**1980**). "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am. **67**, 318-326.

Stephens, S.D.G. (**1976**). "The input for a damaged cochlea - a brief review," Br. J. Aud. **10**, 97-101.

Stevens, S.S. (**1966**). "Power-group transformations under glare, masking and recruitment," J. Acoust. Soc. Am. **39**, 725-735.

Stuart, A. and Phillips, D.P. (**1996**). "Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing and presbyacusic listeners," Ear Hear. **17**, 478-489.

Studebaker, G.A., Sherbecoe, R.L., McDaniel, D.M., and Gwaltney, C.A. (**1999**). "Monosyllabic word recognition at higher-than-normal speech and noise levels," J. Acoust. Soc. Am. **105**, 2431-2444.

Summers, V., and Molis, M.R. (**2004**). "Speech recognition in fluctuating and continuous maskers: effects of hearing loss and presentation level," J. Speech Lang. Hear. Res. **47,** 245-256.

Takahashi, G.A., and Bacon, S.P. (**1992**). "Modulation detection, modulation masking, and speech understanding in noise in the elderly," J. Speech Hear. Res. **35**, 1410-1421.

Ter Keurs, M., Festen, J.M., and Plomp, R. (**1993a**). "Effect of spectral envelope smearing on speech reception. II," J. Acoust. Soc. Am. **93**, 1547-1552.

Ter Keurs, M., Festen, J.M., and Plomp, R. (**1993b**). "Limited resolution of spectral contrast and hearing loss for speech in noise," J. Acoust. Soc. Am. **94**, 1307-1314.

TNS NIPO (**2005**). *Gehoor in Nederland (Hearing in the Netherlands).* Retrieved (20070618) from http://www.nvvs.nl/download/NIPOonderzoek2005.pdf.

Turner, C.W., and Brus, S.L. (**2001**). "Providing low- and mid-frequency speech information to listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **109**, 2999-3006.

Van Rooij, J.C.G.M., Plomp, R., and Orlebeke, J.F. (**1989**). "Auditory and cognitive factors in speech perception by elderly listeners. I: Development of test battery," J. Acoust. Soc. Am. **86,** 1294-1309.

Van Rooij, J.C.G.M., and Plomp, R. (**1990**). "Auditory and cognitive factors in speech perception by elderly listeners. II: Multivariate analysis," J. Acoust. Soc. Am. **88,** 2611-2624.

Van Rooij, J.C.G.M., and Plomp, R. (**1992**). "Auditory and cognitive factors in speech perception by elderly listeners. III: Additional data and final discussion," J. Acoust. Soc. Am. **91,** 1028-1033.

Van Wijngaarden, S.J., Bronkhorst, A.W., Houtgast, T., and Steeneken, H.J. (**2004**). "Using the Speech Transmission Index for predicting non-native speech intelligibility," J. Acoust. Soc. Am. **115**, 1281-1291.

Van Wijngaarden, S.J., and Drullman, R. (**2006**). "Development of a binaural speech transmission index," J. Acoust. Soc. Am. **119**, 3442.

Van Wijngaarden,. S.J., and Houtgast, T. (**2004**). "Effect of talker and speaking style on the Speech Transmission Index", J. Acoust. Soc. Am. **115**, 38-41.

Van Wijngaarden, S.J., Steeneken, H.J., and Houtgast, T. (**2002**). "Quantifying the intelligibility of speech in noise for non-native listeners," J. Acoust. Soc. Am. **111**, 1906-1916.

Versfeld, N.J, Daalder, L., Festen, J.M. and Houtgast, T. (**2000**). "Method for the selection of sentence materials for efficient measurement of speech reception threshold," J. Acoust. Soc. Am. **107**, 1671-1684.

Versfeld, N.J., and Dreschler, W.A. (**2002**). “The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners,” J. Acoust. Soc. Am. **111**, 401-408.

Wagener, K.C., and Brand, T. (**2005**). “Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: influence of measurement procedure and masking parameters,” Int. J. Audiol. **44**, 144-156.

Warren, R.M. (**1970**). “Perceptual restoration of missing speech sounds,” Science **167**, 392-393.

Watson, C.S., Qiu, W.W., Chamberlain, M.M., and Li, X. (**1996**). “Auditory and visual speech perception: confirmation of a modality-independent source of individual differences in speech recognition,” J. Acoust. Soc. Am. **100**, 1153-1162.

Whalen, D.H., Benson, R.R., Richardson, M., Swainson, B., Clark, V.P., Lai, S., Mencl, W.E., Fulbright, R.K., Constable, R.T., and Liberman, A.M. (**2006**). J. Acoust. Soc. Am. **119**, 575-581.

Xu, L., Thompson, C.S., and Pfingst, B.E. (**2005**). “Relative contributions of spectral and temporal cues for phoneme recognition,” J. Acoust. Soc. Am. 3255-3267.

Zekveld, A.A., Deijen, J.B., Goverts, S.T., and Kramer, S.E. (**2007a**). “The relationship between nonverbal cognitive functions and hearing loss,” J. Speech Lang. Hear. Res. **50**, 74-82.

Zekveld, A.A., George, E.L.J., Kramer, S.E., Goverts, S.T., and Houtgast, T. (**2007b**). “The development of the Text Reception Threshold test: a visual analogue of the Speech Reception Threshold test,” J. Speech Lang. Hear. Res. **50**, 576-84.

Zurek, P.M., and Delhorne, L.A. (**1987**). “Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment,” J. Acoust. Soc. Am. **82**, 1548-1559.

Summary

Factors affecting speech reception in fluctuating noise and reverberation

Understanding speech in everyday life can be a serious challenge. One reason is that typical listening situations include a large variety of external distortions, like reverberation or noise. On top of that, speech intelligibility can be seriously hampered by internal distortions, like the presence of a hearing loss. Even when the sound is presented at levels well above the detection threshold, hearing-impaired listeners often complain of not being able to understand what is said. When the background noise level fluctuates over time, as is often the case in daily listening situations, the differences in speech reception between normal-hearing and hearing-impaired listeners appear to be even larger.

The current thesis basically describes a series of experiments, investigating the effects of these internal and external factors on speech intelligibility, after a short general introduction to the subject (Chapter 1). The first part of this thesis (Chapters 2, 3 and 4) assesses the effect of fluctuating noise on speech reception, investigating why hearing-impaired benefit less than normal-hearing listeners from the relatively silent periods in fluctuating noise. Its second part (Chapters 5 and 6) focuses on the effect of reverberation on speech reception, determining whether the relative contributions of noise and reverberation in disturbing speech intelligibility depend on the type of hearing-impairment.

Chapter 2 discusses an experiment in which the Speech Reception Threshold (SRT) in stationary noise and in several amplitude-modulated noises was measured for normal-hearing listeners, sensorineural hearing-impaired listeners, and normal-hearing listeners with simulated hearing loss. The latter group consisted of normal-hearing listeners, who received an additional masking noise, giving rise to an artificially raised hearing threshold. This approach made it possible to determine whether differences in masking release are due to a loss of signal audibility, or due to processing deficits in the ear, such as reduced spectral and temporal resolution. Results show that the reduced masking release can only partly be accounted for by reduced audibility. Instead, auditory temporal resolution and age are shown to be the main factors governing masking release for speech in modulated noise, accounting for more than half of the inter-subject variance. Their influence appears to be related mainly to the higher stimulus frequencies.

It is suggested that the observed adverse effect of age on speech intelligibility in fluctuating noise may be associated with cognitive or other non-auditory factors.

Chapter 3 investigates this suggestion, discussing an experiment in which auditory measurements were performed for normal-hearing and hearing-impairment listeners. In addition, a visual analogue of the SRT test was included in this experiment: the Text Reception Threshold test or TRT. Results show that, for the normal-hearing group, variance in speech reception is mainly associated with non-auditory factors. For hearing-impaired listeners, auditory and nonauditory factors both are shown to be relevant for the comprehension of sentences in fluctuating background noise.

Chapter 4 extends the results of Chapter 3 with measurements in a larger group of age-matched normal-hearing and hearing-impaired participants, also including tests of cognitive non-auditory functions. Results show that the earlier obtained age-effect is indeed related to the importance of non-auditory processing for speech reception in fluctuating noise. In particular, a test of spatial working memory (SWM) and the TRT appear to be relevant for the recognition of speech in fluctuating backgrounds. This finding makes it possible to investigate the sources responsible for deteriorated speech reception in clinical practice.

The effect of reverberation on speech reception is explored in Chapter 5, in which a model is introduced to predict SRTs for normal-hearing listeners for combinations of non-stationary noise and reverberation. The model combines two other models of speech reception, the Extended Speech Intelligibility Index (ESII) and the Speech Transmission Index (STI). First, the effects of reverberation on non-stationary noise (i.e. reduction of masker modulations) and on speech quality are evaluated separately. Subsequently, the ESII is applied to predict the speech reception threshold (SRT) in the masker with reduced modulations. Results show that the model accurately predicts SRTs in non-stationary noise and reverberation for normal-hearing listeners. Moreover, it is shown that the benefit from masker fluctuations may be substantially reduced in listening conditions with reverberation.

Chapter 6 discusses the development of a simple adaptive test, the Speech Reception Reverberation Threshold (SRRT), designed to determine the amount of reverberation that an individual listener can sustain and still understand 50% of the presented sentences. By application of this new test, it is shown that the Speech Transmission Index (STI) is a convenient, single number to quantify speech recognition performance of individual hearing-impaired listeners in listening conditions with noise and reverberation. The nature of the hearing loss (presbycusic, congenital, or otherwise) does not affect this result. It is suggested that determining the effect of reverberation on sentence intelligibility directly, using the SRRT, may be helpful in further understanding the supposed importance of cognitive processing in reverberant listening conditions.

The last chapter of this thesis (Chapter 7) presents a final evaluation. It is concluded that auditory and non-auditory factors both affect speech reception for hearing-impaired listeners, specifically in complex listening situations with fluctuating noise or reverberation. Future research might further investigate the relation between temporal processing and auditory or non-auditory abilities. The relevance of the developed Text Reception Threshold as a tool to determine the non-auditory factors in speech reception is further underlined. Finally, it is suggested that the measurement of speech intelligibility thresholds in complex listening situations is incorporated in standard procedures to clinically assess the challenges of an individual hearing-impaired listener in daily life.

Samenvatting

Het verstaan van spraak in fluctuerend achtergrondlawaai en in galm

In het dagelijks leven kan het behoorlijk lastig zijn om goed te verstaan wat gezegd wordt. Dat komt onder meer doordat er vaak externe stoorzenders aanwezig zijn, zoals galm of achtergrondlawaai. Bovendien wordt het verstaan van spraak soms bemoeilijkt door interne factoren, zoals de aanwezigheid van een gehoorverlies. Zelfs als de spraak zelf hard genoeg is, melden slechthorenden vaak dat ze het wel kunnen horen, maar niet kunnen verstaan. Als het achtergrondlawaai niet constant van niveau is, zoals gebruikelijk in het dagelijks leven, nemen de verschillen tussen normaal- en slechthorenden alleen maar toe.

Dit proefschrift beschrijft een serie onderzoeken naar het effect van deze interne en externe factoren op spraakverstaanbaarheid, na een korte algemene inleiding (Hoofdstuk 1). Het eerste deel van dit proefschrift (Hoofdstuk 2, 3 en 4) gaat over het verstaan van spraak in fluctuerend achtergrondlawaai, ook wel aangeduid als 'ruis'. Er wordt onderzocht hoe het komt dat slechthorenden, in vergelijking met normaalhorenden, minder profiteren van de relatief stille periodes in fluctuerende ruis. Het tweede deel (Hoofdstuk 5 en 6) concentreert zich op het effect van galm op spraakverstaan. Er wordt bepaald of het relatieve belang van ruis en galm bij het verstoren van spraakverstaan afhangt van het type gehoorverlies.

Hoofdstuk 2 beschrijft een experiment, waarin de drempel voor spraakverstaanbaarheid (Speech Reception Threshold of SRT) werd gemeten in stationaire ruis en in verschillende soorten amplitude-gemoduleerde ruis, bij normaalhorenden, slechthorenden en normaalhorenden met een gesimuleerd gehoorverlies. De laatste groep bestond uit normaalhorenden die een extra maskeerruis kregen aangeboden, wat bij hen leidde tot een kunstmatig verhoogde gehoordrempel. Deze aanpak maakte het mogelijk te bepalen of verschillen in spraakverstaan te wijten zijn aan een verminderde hoorbaarheid, of aan verwerkingsproblemen in het oor, zoals een verlies aan spectrale of temporele resolutie. De resultaten laten zien dat verminderd spraakverstaan slechts deels samenhangt met verminderde hoorbaarheid. In plaats daarvan blijken auditieve temporele resolutie en leeftijd de bepalende factoren voor het verstaan van spraak in niet-stationair achtergrondlawaai. Samen verklaren deze twee factoren meer dan de helft van de verschillen tussen proefpersonen. Hun invloed blijkt met name samen te hangen met de verwerking van de hogere geluidsfrequenties.

Het gevonden leeftijdseffect zou samen kunnen hangen met cognitieve of andere niet-auditieve factoren. In Hoofdstuk 3 wordt deze suggestie nader onderzocht, door een experiment waarin niet alleen auditieve metingen werden gedaan, maar waarin daarnaast een visueel analogon van de SRT werd uitgevoerd: de Text Reception Threshold of TRT. De resultaten tonen aan dat spraakverstaan voor normaalhorenden met name samenhangt met niet-auditieve factoren. Voor slechthorenden blijken zowel auditieve als niet-auditieve factoren van belang te zijn bij het verstaan van zinnen in fluctuerend achtergrondlawaai.

In Hoofdstuk 4 worden de resultaten van Hoofdstuk 3 verder uitgebreid met metingen in een grotere groep, leeftijds-gematchte normaal- en slechthorenden, waarbij ook cognitieve niet-auditieve testen werden gemeten. Het eerder gevonden leeftijdseffect blijkt inderdaad samen te hangen met het belang van niet-auditieve verwerking bij het verstaan van spraak in fluctuerende ruis. Met name een werkgeheugen-taak en de TRT blijken relevant voor de spraakverstaanbaarheid. Dit resultaat biedt mogelijkheden om in de klinische praktijk vast te stellen of verminderd spraakverstaan te wijten valt aan auditieve of niet-auditieve factoren.

Het effect van galm op het verstaan van spraak wordt onderzocht in Hoofdstuk 5, waarin een model wordt geïntroduceerd om SRTs te voorspellen voor normaalhorenden in combinaties van niet-stationaire ruis en galm. Het model combineert twee andere spraakverstaanbaarheids-modellen, de recent uitgebreide Spraakverstaanbaarheids-Index (ESII) en de Spraak Transmissie Index (STI). Allereerst worden de effecten van galm op niet-stationaire ruis (het verminderen van ruisfluctuaties) en op de kwaliteit van spraak apart behandeld. Vervolgens wordt de ESII toegepast om een voorspelling te geven van de SRT in de ruis met verminderde modulaties. Het model blijkt inderdaad accurate voorspellingen te geven voor normaalhorenden. Bovendien wordt aangetoond dat het voordeel van ruismodulaties behoorlijk verminderd kan zijn als er galm aanwezig is.

Hoofdstuk 6 bespreekt de ontwikkeling van een simpele adaptieve test, de drempel voor spraakverstaan in galm (SRRT), ontworpen om de hoeveelheid galm te bepalen die een individuele luisteraar kan verdragen om nog juist 50% van de aangeboden zinnen te verstaan. Door deze test toe te passen wordt aangetoond dat de Spraak Transmissie Index (STI) een goede maat is om aan te geven hoe goed een individuele slechthorende nog spraak verstaat in condities met galm en achtergrondlawaai. De aard van het gehoorverlies (ouderdom, congenitaal, of anderszins) blijkt daarbij geen rol te spelen. Door de SRRT kan het effect van galm of de verstaanbaarheid van zinnen heel direct gemeten worden, wat van pas kan komen bij het verder begrijpen van het veronderstelde belang van cognitieve factoren bij het luisteren in galm.

Het laatste hoofdstuk van dit proefschrift (Hoofdstuk 7) is een eindevaluatie. De conclusie is dat zowel auditieve als niet-auditieve factoren voor slechthorenden van belang zijn bij het verstaan van spraak, in het bijzonder in complexe luistersituaties met fluctuerende ruis of galm. Het belang van de ontwikkelde Text Reception Threshold als een mogelijkheid om de niet-auditieve component in het spraakverstaan te bepalen wordt verder onderstreept. Tenslotte wordt voorgesteld dat het meten van het spraakverstaan in complexe luistersituaties wordt opgenomen in standaard klinische procedures. Dat maakt het mogelijk om ook in de klinische praktijk zicht te krijgen op de uitdagingen van een slechthorende in het dagelijks leven.

Dankwoord

Een dankwoord vervult in een proefschrift een bijzondere plek. Het heeft geen vaste omvang of indeling, wordt niet op wetenschappelijke kwaliteit beoordeeld en nergens anders gepubliceerd. Niettemin, of misschien juist daardoor, wordt het vaak als eerste gelezen. Wat mij betreft is dat terecht: belangrijker dan de inhoud van dit proefschrift vind ik de mensen zonder wie dit boekje er nooit geweest zou zijn.

Allereerst zijn dat mijn promotoren, Joost en Tammo, die mij introduceerden binnen de audiologie, nu bijna vijf jaar geleden. Het was mij een genoegen dit project samen met hen tot een goed einde te brengen. Dank voor jullie sturing, jullie creatieve ideeën, en jullie goede feedback (soms tot op de komma).

Daarnaast wil ik mijn collega's 'achter de klapdeuren' bedanken, in het bijzonder: Sophia, voor je enthousiasme en je coaching; Adriana, voor de fijne samenwerking en je nauwgezetheid; Theo, voor je rol als klinische sparring partner, nu en in de toekomst; Janneke, Johannes, Wieneke, en Gaston, voor de pauze's en goed advies als het even allemaal niet wilde; Hans, Jacqueline en Ton, voor de nodige technical support als ik weer eens wat gesloopt had, en de vrolijke (muziek)-noot.

Dan al mijn klinische collega's: jullie hebben mij niet alleen gered van de vierkante ogen, maar zorgden ook voor de jus die het werken in een ziekenhuis zo leuk maakt. Dank jullie wel allemaal voor de afgelopen jaren; ik ben blij dat ik ook de komende tijd met jullie mag samenwerken. En die klapdeuren, die mogen er best nog wel een keertje uit…

Dit onderzoek was niet mogelijk geweest zonder de medewerking van alle proefpersonen, die ik bij deze wil bedanken voor de geïnvesteerde tijd en energie.

Ook buiten mijn werkkring heb ik veel steun gevonden bij alles wat ik deed, of dat nu was in Maastricht, in Eindhoven, in Nijmegen, in Katwijk of in Vleuten. Lieve ouders, familie, vrienden, buren en kennissen, dank jullie wel voor alle support!

De laatste alinea van dit dankwoord is voor mijn vrouw: lieve Anke, dankjewel voor je tomeloze enthousiasme, je inspiratie en je volharding; nu samen op weg naar weer een nieuw avontuur!

Vleuten, 6 september 2007

Curriculum Vitae

Erwin L.J. George werd geboren op 23 november 1979 te Heerlen. Na het voltooien van de middelbare school (VWO) in 1997 verhuisde hij naar Nijmegen, waar hij zich in het studentenleven stortte. Hij leerde de bestuurlijke kant van het Hoger Onderwijs kennen via zijn voorzitterschap van de Facultaire Studentenraad (2000) en zijn lidmaatschap van de VSNU-visitatiecommissie Natuurkunde (2001). In 2002 behaalde hij zijn doctoraal in de Natuurkunde, met een scriptie op het gebied van de toegepaste Hoge Energie Fysica. Via een Masterclass bij TNO en de Koninklijke Marine kwam hij in contact met het Vrije Universiteit Medisch Centrum, waar hij in 2003 een half jaar werkzaam was als audiologie assistent. In juli 2003 startte hij als promovendus binnen de afdeling KNO/Audiologie. In het tweede jaar van zijn promotie (2004) was hij actief binnen het bestuur van het Promovendi Netwerk Nederland. Sinds 1 juli 2007 is hij binnen het Audiologisch Centrum van het VUmc werkzaam als klinisch fysicus audioloog in opleiding.

Erwin L.J. George was born on November 23, 1979 in Heerlen. After finishing his secondary school education (VWO) in 1997, he moved to Nijmegen, where he enjoyed student life. He was introduced to Higher Education policy by chairing the Student's Council of his faculty (2000) and by participating in the VSNU visitation-committee for Physics (2001). He graduated in Physics in 2002, with a thesis in the field of applied High Energy Physics. A Masterclass at TNO and the Royal Dutch Navy led him to the VU University Medical Center, where, in 2003, he worked as an audiology assistant for six months. In July 2003, he started working on this PhD thesis within the department of ENT/Audiology. In his second year as a PhD candidate, he was actively involved as a board member in the PhD Network of the Netherlands. Since July 1, 2007, he has been working as a clinical physicist audiologist (i.o.) at the Audiological Center of the VUmc.